An Iterative Algorithm for Lithography on Three-Dimensional Surfaces

JOSHUA JAMES COWLING

An Iterative Algorithm for Lithography on Three-Dimensional Surfaces

Joshua Cowling

A Thesis presented for the degree of
Doctor of Philosophy

School of Engineering and Computing Sciences
University of Durham
England

April 2015

An Iterative Algorithm for Lithography on Three-Dimensional Surfaces

Joshua Cowling

Submitted for the degree of Doctor of Philosophy
April 2015

Abstract

Photolithography is an optical and chemical process for the patterning of commonly flat substrates with shapes which are useful for electronics and a number of other applications. Holography, in its most general sense, is the manipulation of the coherent properties of an optical wave-front to produce two or three dimensional light patterns. A combination of holography and photolithography therefore allows for the patterning of three dimensional substrates by exploiting the coherence of an optical source. The work in this thesis approaches an optical optimisation methodology centred around the iterative algorithms derived from the Gerchberg-Saxton algorithm. This allows the design of holograms which, in turn, allows the patterning of three dimensional surfaces. The system parameters for the design of two and three dimensional light patterns in this methodology are examined. Simulations and practical optical examples are provided throughout with some application-focused demonstrations. The result is an understanding of an iterative optimisation approach within the context of lithography, and an implementation and methodology for achieving three dimensional patterns.

Declaration

The work in this thesis is based on research carried out at the School of Engineering and Computing Sciences, Durham University, England. No part of this thesis has been submitted elsewhere for any other degree or qualification and it is all my own work unless referenced to the contrary in the text.

Acknowledgements

I would like to acknowledge my colleagues and superiors at Durham University and those we collaborated with at Sheffield University, without whom none of this work would have been possible. Namely: Professor Alan Purvis, Dr Richard McWilliam, Mr Florian Soulard, Dr Gavin Williams, Dr Jesus Toriz-Garcia, Dr Luke Seed and Professor Peter Ivey. I would also like to thank my parents and the great many friends who have supported me throughout this work, as trying as I was at times: Ruth and Paul Michell, Melissa Narraway, Steven Wells, Luke Stanbra, John Rhodes, Alice Blunt, Peff Soulsby and William Thorley. Finally I'd like to thank all of the usual lunch and coffee time gangs who made even the worst days bearable.

Contents

List of Figures

List of Tables

List of Abbreviations

PCB	Printed Circuit Board
MEMS	Micro Electro-Mechanical Systems
2D	Two Dimensional
3D	Three Dimensional
TSV	Through Silicon Via
DOF	Depth Of Focus
OPC	Optical Proximity Correction
RET	Resolution Enhancement Technology
RS	Rayleigh-Sommerfeld
FFT	Discrete Fourier Transform
FFT	Fast Fourier Transform
FDF	Fresnel Diffraction Formula
FC	Fresnel Convolution
AS	Angular Spectrum
CGH	Computer Generated Hologram
SLM	Spatial Light Modulator
CPU	Central Processing Unit
GPU	Graphics Processing Unit
SSE	Sum of Squared Error
GS	Gerchberg-Saxton
S/N	Signal to Noise
IFTA	Iterative Fourier Transform Algorithm
DBS	Direct Binary Search
DMD	Digital Micromirror Device
LCOS	Liquid Crystal on Silicon
PSF	Point Spread Function
NA	Numerical Aperture
FWHM	Full Width at Half Maximum
BW	Bandwidth
COV	Coefficient of Variance
CMOS	Complementary Metal Oxide Semiconductor

Chapter 1

Introduction

Projection lithography has long been the de facto standard for high volume manufacture in the semiconductor industry [1, 2]. Moore's law still holds today and as we broach technology below 10's of nanometres and smaller, it will undoubtedly remain the standard for at least several further technology generations. As a process its success is down to its flexibility and extensibility. Decreasing wavelengths, improved chemical processes and a host of other tweaks and tricks have evolved this technology to its current state, and this shows no sign of stopping.

These current processes, however, are not without limitation. Systems with more complex requirements are emerging. The need for the close integration of both regular silicon and high speed gallium-arsenide (and other) substrate devices as well as standard wire-bond package and PCB connection systems can become unwieldy, problematic and expensive in modern system and package design. Larger scale Micro Electro-Mechanical Systems (MEMS) may be built from non-planar substrates for which conventional mask design begins to become inadequate. Furthermore, as the number of devices on a single chip increases, focus turns to 3D or chip-stacking methods to obtain the required interconnection density.

This project, in conjunction with other research being carried out at Durham and Sheffield Universities, aims to bring a practical design strategy to the implementation of 'holographic' lithographic masks. This thesis specifically develops the *optical* design methodology required to design diffractive systems which allow the 3D patterning of substrates in an attempt to break down the planar constraints of conventional lithography.

Though it is not the aim of this thesis to fully realise a lithographic solution for a multi-billion dollar semiconductor industry, it is thought that this technology and approach could greatly help with the problems described above by further extending the lithographic process to cope with non-planar substrates.

To this end, we will deal initially with implementing a computational optical model of sufficient accuracy which - when combined with the iterative optimisation algorithms discussed later in this thesis - will allow the design of 3D light fields which we may then apply to lithography. Having developed and implemented an iterative optimisation method, we will then characterise some of its properties regarding the parameters of simulation and implementation, and begin to define the capabilities of the algorithm to design holograms useful for lithography. As we proceed, modifications to the algorithm will be introduced to solve specific issues and the effects of these modifications will be investigated.

1.1 Thesis Structure

- Chapter 1 provides a brief introduction and some background discussion on optics, lithography and holography.
- Chapter 2 is a review of derivations for some fundamental optical theory, including physical interpretations to clarify certain aspects. Furthermore this chapter explores approaches to modelling diffraction and touches on the issues of implementation and computation.
- Chapter 3 reviews existing iterative optimisation algorithms and explores some of their implementations and limitations.
- Chapter 4 describes the practical implementation of a holographic optical system and discusses the various benefits and limitations imposed by these practicalities.
- Chapter 5 looks at some of the fundamental limits of optical simulation as they apply to our chosen methods and application and then discusses the implementation and results for a planar iterative optimisation scheme. It then goes on to examine some known limitations and our approach for overcoming them (Error reduction, multi-step propagations, filtering and pixelation, image localisation, feature size, background value, noise suppression and coherence).
- Chapter 6 generalises the work of chapter 5 to a method that works for images in 3D and addresses the problems incurred as a result of this generalisation. It includes analysis of the expected and actual output of the modifications which allow 3D fields to be generated.

- Chapter 7 shows three implementations of our 3D patterning method, and investigates and discusses some further practical issues encountered in the production of these application-led patterns.

- Chapter 8 concludes our research, and discusses the future of the project.

Practical implementation is discussed throughout this thesis, but implementation details such as code examples are mostly confined to the appendix for brevity. Appendix A gives some commented Matlab implementations for analysis and propagation schemes.

1.2 History of Optics and Holography

Science and philosophy have struggled to come to terms with the nature of light since the time of the Ancient Greeks. From the point of view of modern science, Dutch mathematician and physicist Christiaan Huygens was among the first to formalise the wave properties of light in the 17th century. His ideas would be challenged by the work of Sir Isaac Newton, who explained the basic properties of light as a particle. The interference experiments of Thomas Young would go some distance to proving light's wave nature, but as we now know (from the work of many of the great scientists of the recent past), all matter can exhibit both particle and wave properties.

From the 18th century to the current day, much work has gone in to finding suitable simple mathematical relationships describing the wave propagation of light. Augustine-Jean Fresnel is associated with probably the most famous approximation, which we consider later.

The Hungarian-born Dennis Gabor invented the techniques by which phase information could be simply stored on photographic media in 1947 [3]. Initially intended to improve X-ray and electron microscopy applications, it was not until the invention of the laser that Gabor's photographic technique - now known as holography - could become more widely appreciated.

Modern holography is a wide and diverse area covering physical and digital recording (e.g. digital holographic microscopy), as well as real and computer generated synthesis of patterns, (e.g. 3D display and holographic optical tweezers).

1.3 An Introduction to Holography

The basic holographic process involves taking a coherent light source and illuminating an object. The light reflected and scattered from the object then illuminates a photosensitive screen. We call this the object beam. Superimposed on the screen is the reference beam; a plane wave illumination which is split away from the source prior to illumination of the object. Assuming that both object and reference beams interfere coherently (which is not a simple assumption to make), the interference between the two waves encodes information about the phase of the wave in the intensity pattern recorded on the screen.

Without a reference wave the intensity is described by

$$I_{\text{screen}} = |U_O|^2 \tag{1.1}$$

With a reference wave however, because the intensity of the wave is the square of the sum of the superimposed fields, the intensity is described by

$$I_{\text{screen}} = U_O U_R^* + |U_R|^2 + |U_O|^2 + U_O^* U_R \tag{1.2}$$

where U_O is a 2D complex field describing the object beam, and U_R is the reference beam. The terms of interest here are the two terms in which the object and reference beams interact. They represent information about the object field which would not be present if only measuring the magnitude of the object beam. When the hologram is replayed by illuminating the recorded hologram with a reference beam but without the object beam, the field of the reference wave is modulated by the previously recorded intensity pattern, which contains information about the relative phases of the object and reference beams in the modulation pattern. In comparison a recording without the reference wave results only in information about the intensity of the object wave being recorded.

The result of the inclusion of this phase information is that when the hologram is replayed the now absent object is visible as both real and virtual images through the hologram because of the modulation of the reference wave. Light now converges to a focal image not in the plane of the hologram modulation, as though the object were still in place. This gives a three-dimensional impression to an observer. Causing out-of-plane manipulation of the convergence of light is the effect which we will exploit to form images on surfaces which are not flat.

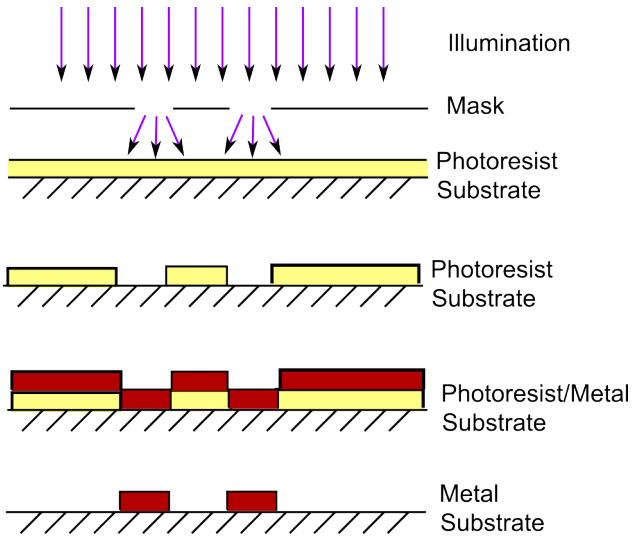


Figure 1.1: Simple photo-lithographic process

1.4 An Introduction to 3D Lithography

The term "Lithography" describes the printing process used to produce modern electronic patterns on a variety of scales. More accurately termed photolithography, the process (which is denoted diagrammatically in figure 1.1) involves taking a nominally flat surface, spin-coating it with a photosensitive polymer, and then selectively exposing parts of the polymer to light by illuminating it through a mask pattern. This layer is altered by the illuminating radiation such that when chemically processed the exposed photoresist is either crosslinked so that it remains whilst other unexposed areas are removed (positive acting photoresist), or un-crosslinked so that they are removed whilst the unexposed areas remain (negative acting photoresist).

Recent high profile developments have begun to require electronic patterns with larger depth components, but these methods work within the limitation of existing photo-lithographic processes. Micro Electro-Mechanical Systems (MEMS) development involves the creation of tiny mechanical structures, but often takes on the standard layered monolithic development process [4] of lithography. Using post-processing to achieve etching below some elements, this results in a 3D structure built from sequential layers and processing. On a smaller scale, Intel's new transistor level technologies tout "3D" structures, [5] which in reality are thinner, deeper elements and not *significantly* non-planar in a larger context. On a larger scale, we encounter the need to stack wafers or packages for greater levels of interconnection between components. In many of these cases the processes used are more difficult than standard lithography and require special manipulation and handling of devices to integrate correctly. For example, "Through Silicon Via" (TSV) connections usu-

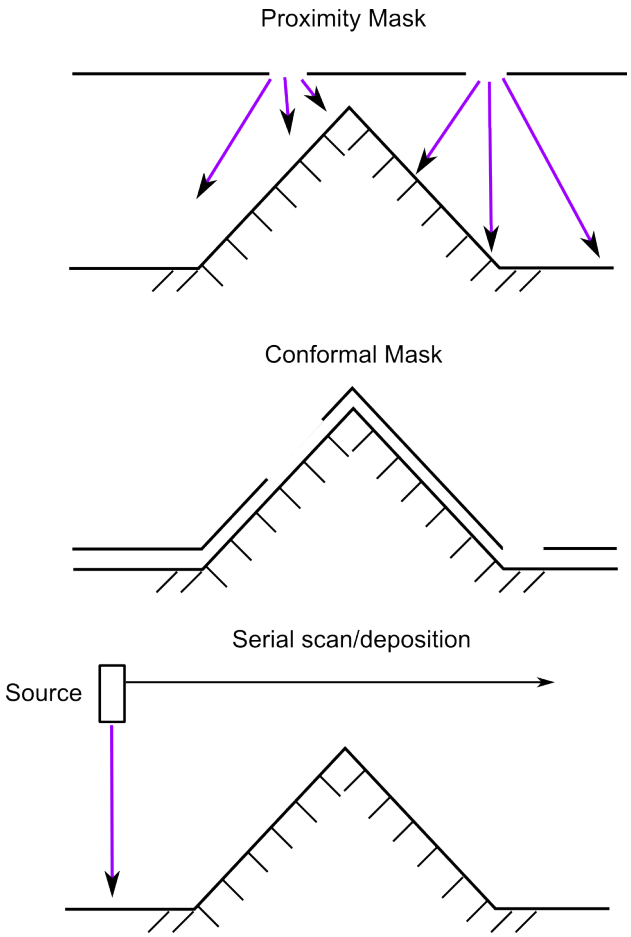


Figure 1.2: Approaches to 3D lithography

ally require thinning of the chip, which can lead to fragility and failure. Connections between stacked chips require layers of many wire bonds, a process which is slow and error-prone. The potential is clear; an extension allowing the advantages of a standard lithography process effectively applied to such 3D structures is a useful step for future electronics integration and packaging.

Some simple approaches to lithography or similar electronic patterning in 3D do already exist, some of which are shown in figure 1.2. These, however, are almost all limited to serial processes, such as laser ablation or serial laser exposure, where an intense and small spot is scanned across a surface to form a pattern. Serial deposition is another approach in which a write head squirts conductive material onto a surface, again limited by speed. Stereo-lithography is a 3D manufacture process in which a 3D model is created from laser-scanned layers of photosensitive polymer in a bath. Ideally a photo-lithographic process is much faster than this kind of serial process - and much more suited to scaling with increased manufacture volume - and can produce patterns with a single exposure.

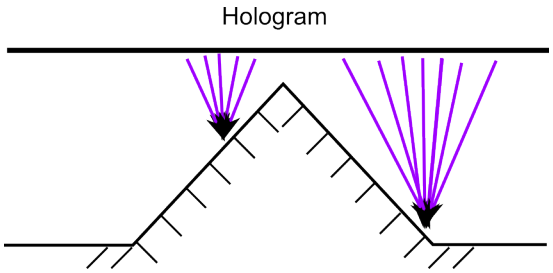


Figure 1.3: Holographic lithography

1.4.1 The Limitations of Conventional Lithography

Conventional lithography utilises either projection or proximity lithography. For proximity, the projection gap limits the resolution of the achievable pattern as diffraction occurs across the gap between mask and substrate. In projection lithography the position of the focal image can be controlled by the position and focal length of the projection lens, but unless a 3D mask is used this focal image is only useful over a short range of z within the depth of focus (DOF) of the optical system. This usable depth can be quantified as the "exposure latitude" [6]. Previously, Kersten [7] has explored the limitations of conventional proximity mask lithography and found simple expressions for the limitations on resolution for a given gap.

Projection lithography is generally used for state-of-the-art nanometre resolution patterning. This involves a set of photo-reduction optics [8] to shrink a mask pattern which may illuminate either a section of a die in a step and repeat (stepper) configuration, or a whole wafer. Illumination systems for these cutting edge lithography systems use short wavelength Excimer lasers and carefully tune coherence effects to achieve maximum resolution. Proximity lithography is a simpler and more cost-effective method, but is generally limited to the micrometer scale or larger. In projection configurations much research has been done into the application of various mask structures to improve the resolution obtainable by modification of mask patterns. Examples of this are "Resolution Enhancement Technology" (RET) wherein diffractive phase-shifting elements are applied in mask patterns and "Optical Proximity Correction" (OPC) elements where shaped elements beyond the standard mask pattern are used to adjust diffractive shapes favourably. In proximity configurations, however, such approaches have largely been ignored as generally such high tolerance patterns are not required. There are, however, several cases where such basic limitations become a serious problem. These cases are largely caused by problematic substrates where a gap is required between mask and substrate.

The application of Holographic lithography (figure 1.3) solves this projection gap issue by forming a focal image at some distance from the hologram plane. Moreover,

the focal image need not be 2D. 3D focal patterns open a wide variety of new substrate and interconnection possibilities to next generation electronics manufacture. It is for these reasons that 3D holographic lithography is likely to be an interesting and useful extension to lithographic processes. In addition, holograms have other intrinsic advantages such as an inherent tolerance to small defects in mask production due to the distributed nature of the information in the hologram.

1.5 Published work associated with this thesis

This document is the culmination of four years' work. During this time, several papers have been published by myself and my colleagues which provide a good outline of the methodology, applications and boundaries of the project and supplement the work in this thesis. For the interested reader, these publications in chronological order are:

- Patent: Improvements in or Relating to Holography [9] : Overview of 3D iterative holographic lithography.
- Journal: Three-Dimensional Holographic Lithography by an Iterative Algorithm [10] : Cowling et al. Published in Optics letters in 2011. Outlines an iterative algorithm to design phase only holograms which uses multiple constraint planes. This algorithm is then applied to produce a dense bus of lines on a sloped surface.
- Journal: Vertical tracks on the side wall of a silicon die using 3D holographic photo-lithography [11] : Toriz-Garcia et al. Published in 2011 in the Journal of Micromechanics and Microengineering. Describes a methodology for applying phase only line holograms for lithography on vertical surfaces and over right angle bends which has been tested on cut silicon die samples.
- Conference: Time-Division Multiplexing of Iterated Holograms for Lithography on 3D Surfaces [12] : Cowling et al. Conference paper published 2012 at OSA: Digital Holography and Three-Dimensional imaging conference in Miami, Florida. Describes a methodology for using active modulators to extend the patterning ability of phase only holograms in 3D.
- Conference: Holographic Lithography on Vertical Surfaces [13] : Conference paper published 2013 at OSA: Digital Holography and Three-Dimensional imaging conference. Describes a methodology for vertical patterning using

active modulators and an iterative algorithm which allows for increased image complexity over analytical patterns.

- Journal: Fabrication of a 3D electrically small antenna using holographic photo-lithography [14] : Toriz-Garcia et al. Published 2013 in the Journal of Micromechanics and Microengineering. Describes the design and optimisation of a hemispherical antenna produced using 3D holographic lithography.

1.6 Declaration of individual work

The research around this thesis was performed by a very talented group of researchers at Durham and Sheffield universities as outlined in the papers above. The work documented in this thesis, however, is all my own except where explicitly referenced or stated. All simulations were based on my own code built in Matlab and all experimental data were experimentally recorded on an optical set up outlined in Chapter 4 built and tested by myself at Durham university. The work of chapter 7 outlines several published applications examples that were produced with the aid of several researchers at Sheffield university who are appropriately referenced in that chapter, and the experimental exposure and development on photoresist for those cases were either produced entirely at Sheffield or with the the aid of myself whilst I visited and worked at Sheffield university.

Chapter 2

Review of Coherent Diffraction Theory

2.1 Introduction

Initially it is important to discuss and consider the methods used to compute optical propagations in free space and through a lens. This is for several reasons. Firstly, these processes will be required to analyse our newly designed holograms. Secondly, the optimisation algorithms we later discuss will have these methods at their core. Thirdly, understanding these methods will give us some context for the sampling constraints required for accurate simulation and optimisation and thus some understanding of of the computational complexity of these derived algorithms.

This chapter begins by showing the derivation of some common and useful optical propagation equations derived from the Huygens-Fresnel principle and the Rayleigh-Sommerfeld approximation. It then goes on to discuss their validity and constraints in terms of simulation and also to define their performance in a numerical analysis context. Some extensions to these methods of numerical simulations are touched upon (rotational transformations, and apertures approximations), before concluding with demonstrations of a selection of phase-only analytical holograms and their limitations in terms of lithographic processing.

2.2 Derivation of Propagation Equations

2.2.1 Rayleigh Sommerfeld Solution

The Huygens-Fresnel principle (figure 2.1) states that each point on a wave-front can be regarded as the source of a new disturbance which, when further propagated

and superimposed, forms the advancing wave front. Though this is not a complete representation of the wave mechanics of light it has proven to be a valuable approximation in understanding its properties, and in many cases is sufficient to predict experimental results. Goodman [15] derives the Rayleigh-Sommerfeld solution to

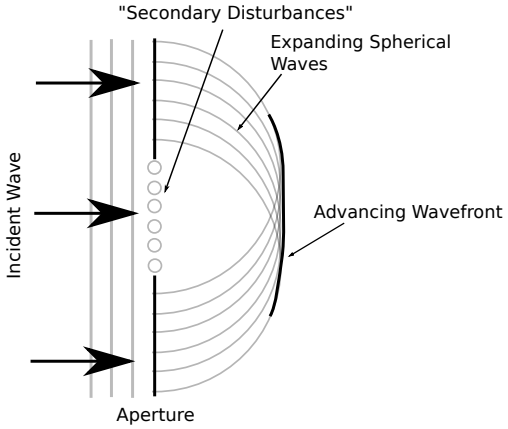


Figure 2.1: The Huygens-Fresnel principle illustrated graphically.

the scalar wave equation as predicted by the Huygens-Fresnel principle as:

$$V(x,y) = \frac{1}{j\lambda}\iint U(\xi,\eta)\frac{e^{jkr}}{r}cos(\theta)d\xi d\eta \tag{2.1}$$

where

$$r(x,y,\xi,\eta) = \sqrt{(x-\xi)^2 + (y-\eta)^2 + z^2} \tag{2.2}$$

In which r is the distance between source and image points separated by a distance z along the optical axis as figure 2.2. $V(x,y)$ is the field in the output plane (coordinates x and y), $U(\xi,\eta)$ field in the input plane (coordinates ξ and η). $j = \sqrt{-1}$, λ is the wavelength of the monochromatic light field. $k = 2\pi/\lambda$ and θ is angle from the optical axis.

A physical interpretation of the kernel of this integration can be seen as summing across spherical waves emanating from all points on a diffractive aperture or object to get the value of the field at a single point. This is augmented by terms $1/j\lambda$ and $cos(\theta)$. Physically these terms are harder to understand, though the cos(θ) term implies a probabalistic directionality that is otherwise absent.

2.2.2 Fresnel Approximation

Reduction of the Rayleigh-Sommerfeld equation allows us to break it down into a much more friendly and usable form. To do this we first constrain our system to two parallel planes centred along the optical axis, illuminated by a plane wave. From

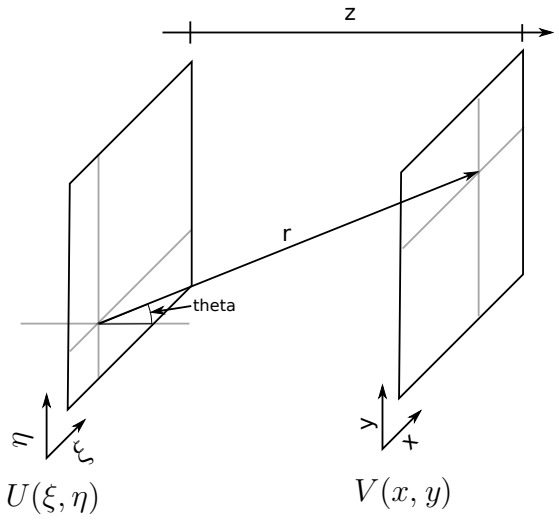


Figure 2.2: r vectors of the Rayleigh-Sommerfeld integral

Fig 2.2 It is trivial to see that:

$$cos(\theta) = \frac{z}{r} \tag{2.3}$$

Substituting this back into 2.1:

$$V(x,y) = \frac{z}{j\lambda} \iint U(\xi,\eta) \frac{e^{jkr}}{r^2} d\xi d\eta \tag{2.4}$$

To further simplify the transform, it is now necessary to make approximations to the r term.

By binomial expansion of a square root such as:

$$\sqrt{1+a} = 1 + \frac{1}{2}a + \frac{1}{8}a^2 + \dots \tag{2.5}$$

It can be seen that factoring z outside and taking the first 2 terms of the binomial expansion of r we are left with:

$$r \approx z\left(1 + \frac{1}{2}\left(\frac{x-\xi}{z}\right)^2 + \frac{1}{2}\left(\frac{y-\eta}{z}\right)^2\right) \tag{2.6}$$

For the r^2 in the denominator of 2.4 the error introduced by taking just the first term, $r = z$, is relatively small if we are considering paraxial systems with $z > x$ or y. This is however not the case for the term in the exponent which is multiplied by a very large number ($k = 2\pi/\lambda$). It is therefore usually pertinent to take two terms from the expansion for this r value (as stated by Goodman [15]), and thus 2.4

becomes:

$$V(x,y) = \frac{1}{j\lambda z}\iint U(\xi,\eta) exp\left(jk\left(z + \frac{(x-\xi)^2}{2z} + \frac{(y-\eta)^2}{2z}\right)\right) d\xi d\eta \tag{2.7}$$

Now taking a factor of e^{jkz} outside of the integration:

$$V(x,y) = \frac{e^{jkz}}{j\lambda z}\iint U(\xi,\eta) exp\left(\frac{jk}{2z}\left((x-\xi)^2 + (y-\eta)^2\right)\right) d\xi d\eta \tag{2.8}$$

From here the formula may be simplified further to give a Fourier transform kernel, or re-arranged to give a convolution. Both arrangements are interesting to consider for their properties under analytical and numerical evaluation.

2.2.3 Fresnel Diffraction Formula

Further factorisation leads to:

$$V(x,y) = \frac{e^{jkz}}{j\lambda z} e^{\frac{jk}{2z}(x^2+y^2)} \iint \left[U(\xi,\eta) e^{\frac{jk}{2z}(\xi^2+\eta^2)}\right] e^{-j\frac{k}{z}(x\xi+y\eta)} d\xi d\eta \tag{2.9}$$

It is important to note here that this formulation is that of a Fourier transform with pre- and post-multiplication by quadratic phase factors with a scaling constant. This step allows us to evaluate simple aperture functions with far smaller computational burden by exploiting the Cooley-Tukey [16] or similarly derived FFT algorithm which can evaluate a Fourier transform in computations of the order $O(N\log(N))$ - as opposed to a naive algorithm which might take up to $O(N^2)$.

Though this Fresnel form lends itself to analytical solution by the use of an analytical Fourier transforms, and is the preferred approach for the physicist attempting to analyse the properties of diffraction for a given situation [15], this arrangement can be practically useless for the engineer attempting to apply modelling of a given application. This is because of the scaling and sampling limitations as discussed in section 2.5.1. Therefore we look to other forms for numerical modelling applications.

2.2.4 Fresnel Convolution

Returning to equation 2.8 simple rearrangement and substitution yields:

$$V(x,y) = \iint U(\xi,\eta) h(x-\xi, y-\eta) d\xi d\eta \tag{2.10}$$

where

$$h(x,y) = \frac{e^{jkz}}{j\lambda z} e^{\frac{jk}{2z}(x^2+y^2)} \tag{2.11}$$

Here we now have a convolution of the input modulated plane, with an impulse response of free space propagation.

By convolution theorem, this can also be evaluated by Fourier transform. Symbolically:

$$\mathcal{F}[V(x,y)] = \mathcal{F}[U(\xi,\eta)]\mathcal{H}(\nu_x,\nu_y) \tag{2.12}$$

where

$$\mathcal{H}(\nu_x,\nu_y) = \mathcal{F}[h(x,y)] = e^{jkz}e^{-j\pi\lambda z(\nu_x^2+\nu_y^2)} \tag{2.13}$$

where $\mathcal{F}[]$ represents the Fourier transformation and ν_x, ν_y are coordinates in the 2D "spatial frequency" domain. This form lends itself much more neatly to numerical analysis, though it is less efficient than the Fresnel transform (see section 2.5.3).

2.2.5 Fraunhofer Approximation

Making a further approximation to the Fresnel Diffraction Formula (FDF) in 2.9, it is possible to simplify this equation to almost a single operation. If z is large in proportion to the size of the diffracting aperture, specifically $z \gg (\xi^2+\eta^2)$, then the phase term on the inside of the Fourier transform, $e^{\frac{jk}{2z}(\xi^2+\eta^2)}$, becomes negligible.

Furthermore, in the simple forward propagation case, we may be uninterested in the phase of the resultant pattern, since intensity is the only measurable quantity we can derive from simple measurement apparatus. In this case, the phase factor outside of the Fourier transform $e^{\frac{jk}{2z}(x^2+y^2)}$ may also be ignored. Finally the scaling factor, $\frac{e^{jkz}}{j\lambda z}$ may also be neglected if we are only interested in the pattern produced rather than the absolute intensity measurement.

All of this leads to the simple transform:

$$V(x,y) = \iint U(\xi,\eta)e^{-j\frac{k}{z}(x\xi+y\eta)}d\xi d\eta \tag{2.14}$$

Which is simply the Fourier transform of the input plane, considering a transform of coordinates as described by the numerical evaluation parameters.

2.3 Three-Dimensional Fields

As hinted at by the the simplification of the diffraction equations to the Fraunhofer approximation, the spatial frequency representation of a pattern, i.e. its Fourier transform, and its diffraction pattern are closely linked.

Singer et al [17], Piestun et al [18] and Whyte et al [19], all approach the understanding of diffraction theory and scalar wave propagation from the perspective of

3D fields.

Singer et al [17] specifically, takes the Fourier transform of a 3D field we obtain its "angular spectrum". In the same way that a time dependent oscillatory sequence can be said to contain waves of the frequencies of its Fourier transform, a 3D field can be composed of the plane waves described by its 3D Fourier transform. Each point in the angular spectrum represents a plane wave travelling in a specific direction with a given amplitude and phase.

For a monochromatic scalar field, the frequency space representation of the field must be confined to a a surface which obeys the Ewald equation.

$$|k| = \sqrt{k_x^2 + k_y^2 + k_z^2} = \frac{n2\pi}{\lambda} \tag{2.15}$$

Where k is the wave vector that describes an arbitrary plane wave. This can also be written in terms of spatial frequency using the identity $2\pi\nu = k$

This equation then represents a sphere of radius n/λ in the frequency domain and can be shown to be the frequency domain counterpart of the Helmholtz wave equation.

2.3.1 Angular Spectrum Propagation

Using the AS representation we can derive another convolution form which retains the merits of the Fresnel Convolution form but does not make any of the approximations. This is derived mathematically in [17] but also has a clear physical description in the spectral domian.

This may be expressed as:

$$V(x, y) = \mathcal{F}(U(\xi, \eta)) \times \mathcal{F}(h(x - \xi, y - \eta)) \tag{2.16}$$

where

$$\mathcal{F}(h(x, y)) = H(\nu_x, \nu_y) = e^{j2\pi z\sqrt{(1/\lambda)^2 + \nu_x^2 + \nu_y^2}} \tag{2.17}$$

Regarding the plane wave spectra of the same 3D wave on two planes, and assuming that no modulation has occurred between the two planes, we can devise the AS kernel by considering the modification of each plane wave between the two planes. The only difference between the spectra is the rotation of each of the plane wave phases by a fixed factor, dependant upon the angle of the wave, which is exactly the function described by the kernel shown above. No other assumptions or modifications need to be made to directly derive this equation from the physical situation.

It is possible to show that the Fresnel convolution is an approximation to this transfer function [17]. However, evaluation of this form is slightly more complex. A limit on the frequencies here must be considered, as frequencies beyond this limit may represent evanescent waves (near field non-radiating waves) in a similar vein to points which exist off the surface of the Ewald sphere.

Very little theoretical and practical difference exists between the Fresnel Convolution and Angular Spectrum approaches, computationally and mathematically. The Angular Spectrum approach uses less approximation to the distance terms of the Rayleigh-Sommerfeld equation, and therefore more appropriate for short propagation distances where transfer function has high frequency in the spectral domain. This is potentially important for later in this thesis where repeated short propagations are applied to constrain an image volume.

2.4 Propagation Validity

Because of the approximations inherent in many of the transforms, as well as the methods of evaluation, certain conditions must be met to ensure their validity when simulating a propagation.

- Firstly, the approximations assumed within the analytical derivations must be adhered to. This means that of scalar theory inherently as well as the approximations to the r vector for Fresnel, Fraunhofer and Angular Spectrum and their far/near field relationship.
- Secondly it must be ensured that the phase and transfer functions to be used are sampled sufficiently such that they do not alias when numerically evaluated.
- Thirdly, since we will be evaluating Fourier transforms and convolutions which have been discretised and evaluated with a fast Fourier transform, it must be assumed that the spatial function that describes the diffraction or image plane is periodic in all space. To stop neighbouring images from interfering in the transform, the function must be embedded in an image of zeros of sufficient width and subsequent repropagation must discard invalid information outside of this range [20].

2.4.1 Analytical Approximations

The RS integral makes only those approximations pertinent to scalar wave theory. This can be accounted for by ignoring polarization effects (simply polarize the

source to be used), by ensuring that the diffracting aperture is much larger than the wavelength, and by not observing the field too close to the diffracting aperture (Goodman, [15] page 35). The Angular Spectrum is the closest to the RS integral (taking the fewest approximations) and therefore is the most accurate approach.

The accuracy of the Fresnel transform (and its associated convolution) is limited by the approximation made to the r vector in its derivation. Though technically this limits its use to quite a large propagation distance behind the aperture (25cm for a 1cm aperture at $\lambda = 0.5\mu m$) as again discussed by [15] (page 69), the transform produces valid results for fields at much shorter distances than this.

The Fraunhofer approximation making yet further approximations to the propagation distance can be shown to be valid only when the Fresnel number is very small, i.e.

$$F \ll \frac{a^2}{D\lambda} \tag{2.18}$$

Where F is Fresnel number, a is aperture dimension, D is propagation distance and λ is wavelength.

2.4.2 Numerical Aliasing

To address the aliasing of analytical functions inherent within these numerical transforms we can take a local frequency approach to define the frequency of a phase function at a given point, dependent upon its absolute position in space (or frequency space). Since the transfer and phase functions of the optical transforms are circularly symmetric we need only do this in one dimension and consider the maximum distance from the zero in our sampled space. This can be compared to the worst case sampling criterion (i.e. the nyquist rate of the simulation) to calculate the minimum or maximum distance at which a transfer function or phase factor will begin to alias and thus introduce noise.

To calculate the frequency of a phase function at a point in either normal or spectral space we use the local spatial frequency approximation given by Goodman [15]:

$$f_l(x) = \frac{\partial}{\partial x}\frac{\phi(x)}{2\pi} \tag{2.19}$$

Where ϕ is the phase of the function in question i.e.

$$F(x) = e^{j\phi(x)} \tag{2.20}$$

This results in the following set of local spatial frequency functions for each method considered:

Method	Related Function	Local Frequency $f_l(x)$ or $f_l(f_x)$
Fresnel Convolution	$\mathcal{H} = e^{jkz} e^{\frac{-j\pi}{\lambda z}(\nu_x^2+\nu_y^2)}$	$\lambda z f_x$
Angular Spectrum	$\mathcal{H} = e^{j2\pi z\sqrt{(1/\lambda)^2+\nu_x^2+\nu_y^2}}$	$(z f_x)/(\sqrt{(1/\lambda)^2 - f_x^2})$
Standard Fresnel	$e^{\frac{jk}{2z}(\xi^2+\eta^2)}$	$(k f_x)/(2\pi z)$
Simple Fraunhofer	None	No transform phase term

It can be seen from the values of local frequency in these functions that for a fixed sampling scheme (number of samples and sample pitch) in space or spatial frequency space the local frequency will change with z. In the convolution methods (Fresnel convolution and Angular Spectrum) the evaluation of these functions are done in the frequency domain, whereas for the standard Fresnel diffraction transform this occurs in a spatial domain. As can be seen from these derived functions the convolution approaches have an increased local frequency for longer z (i.e. z is on top of the function whereas in the Fresnel transform the spatial frequency actually decreases with z). When we introduce the constraints of Shannon-Nyquist sampling theorem i.e.

$$max(f_l) \leq \frac{f_s}{2} \tag{2.21}$$

Where $max(f_l)$ is the local frequency at the furthest extreme of the sampled function, and f_s is the sample pitch of the numerical evaluation. This relation can therefore be used to determine whether a transfer function is well sampled given a sampling regime and propagation distance. Examples of well- and poorly-sampled transfer functions can be seen in 2.3.

We discuss a propagation aliasing distance limit, its physical consequences and its effect on our derived iterative algorithm in section 5.7.1.

2.4.3 Other Factors

Other limits also restrain the usefulness of each of these methods. All above approaches ignore the existence of the so-called "evanescent" waves that occur as standing waves around diffracting apertures. These waves however carry no energy away from the aperture and are localised to only a few wavelengths beyond and behind it, though they can be interacted with for useful effect in certain cases [21].

Finally the sampling must be considered high enough to represent the image pattern in the diffraction plane, and as such respect standard sampling theorem. This is not quite the trivial case of merely correctly representing an expected image

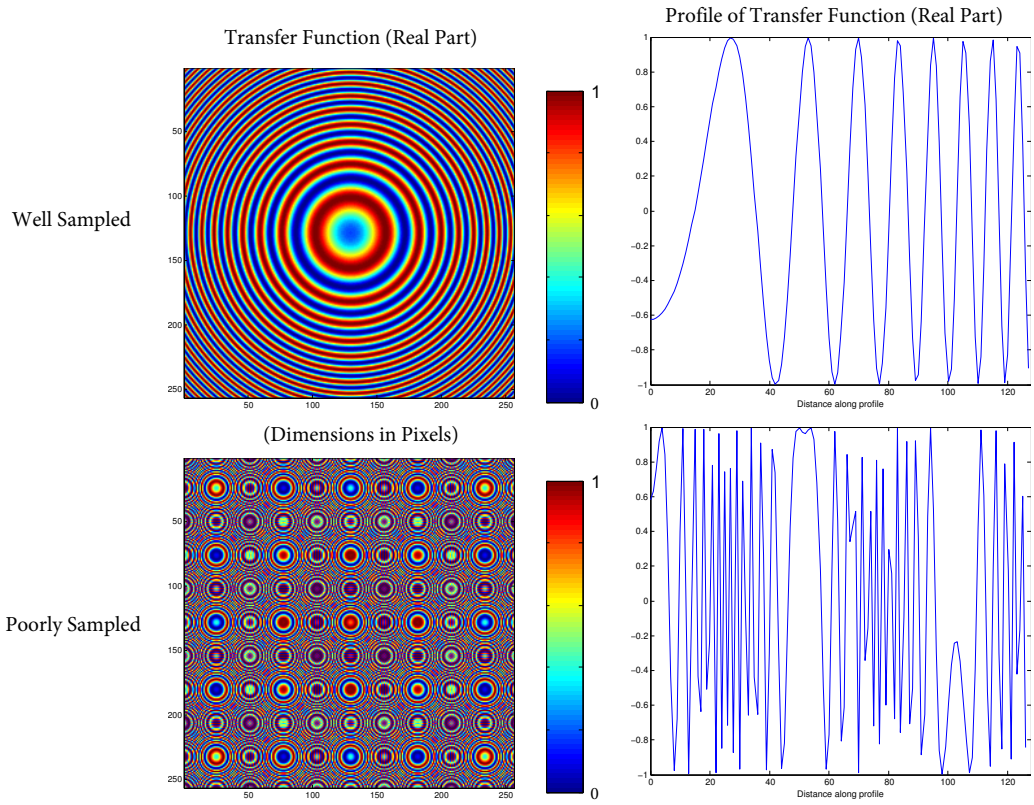


Figure 2.3: Well and poorly sampled transfer functions (real part) for the Fresnel convolution approach

pattern and implementation issues with regard this implementation and the iterative algorithms applies will be discussed as appropriate in section 5.8.1.

2.5 Numerical Evaluation

The details of the computational implementation of this mathematics is of serious concern. Whilst mathematical form may be acceptable for exploiting the analytical solutions in some cases, it becomes incredibly important when considering numerical evaluations in both terms of accuracy and efficiency. This is especially true when dealing with the iterative methods to be discussed later, as sub-sampling a pattern and performing iterative optimisation will reinforce erroneous and noisy patterns, severely degrading image quality.

2.5.1 Fresnel and Fraunhofer Diffraction Formulae

Starting with the FDF 2.9 and comparing with the standard Fourier transform:

$$V(x,y) = \frac{e^{jkz}}{j\lambda z} e^{\frac{jk}{2z}(x^2+y^2)} \iint \left[U(\xi,\eta) e^{\frac{jk}{2z}(\xi^2+\eta^2)} \right] e^{-j\frac{2\pi}{\lambda z}(x\xi+y\eta)} d\xi d\eta \tag{2.22}$$

$$F(\nu_x,\nu_y) = \iint \left[f(\xi,\eta) \right] e^{-j2\pi(\nu_x\xi+\nu_y\eta)} d\xi d\eta \tag{2.23}$$

We see here that a constant term $1/\lambda z$ is included in the Fourier exponent in the Fresnel transform. This term *cannot* be neglected. If we make the substitution $\nu_x = z\lambda x$ and $\nu_y = z\lambda y$ then we can see that the above equations are of the same form. As a consequence, output plane coordinates must be multiplied by a factor of λz to match transform coordinates.

In addition to this, when discretising the FT into the DFT an additional factor of $1/N$ is introduced where N is the total number of samples operated upon in one dimension. For simplicity we will always consider an $N \times N$ grid. In 2D image terms this becomes $1/K$ where K is the total *width* of recording i.e. $K = N\delta$ where delta is the sample pitch on $U[x,y]$.

These two sample scalings therefore combine to relate input an output sample spacings for a numerically evaluated FDF transform such that

$$\delta_V = \frac{\lambda z}{N\delta_U} \tag{2.24}$$

Where δ_V is the sample spacing on the output plane and δ_U is that of the input plane.

The discretisation into the DFT and subsequent evaluation by FFT also brings with it other troubles inherent with this form. The image plane must be assumed to be periodic; padding and sample pitch have an inverse effect on output transform.

2.5.2 Convolution based Transforms (Angular Spectrum and Fresnel Convolution)

Transforms evaluated as a convolution need not suffer from this scaling of input and output sample spacings in the same way that the Fourier transform methods do. The convolution methods are evaluated as follows:

$$V(x,y) = \mathcal{F}^{-1}\left[\mathcal{F}[U(\xi,\eta)]\mathcal{H}(\nu_x,\nu_y)\right] \tag{2.25}$$

Where $\mathcal{H}(x,y)$ is the transfer function of the appropriate convolution method.

Because both forward and inverse Fourier transforms are employed to evaluate the numerical convolution, efficiency is reduced but sample spacing remains constant between input and output planes.

2.5.3 Computational Efficiency

In the case of the computationally intensive operations of iteration which are to be investigated, it is the efficiency of the individual propagation transform that is of utmost importance, though other factors are also significant. The number of floating point operations evaluated for each transform approach is summarised here

Method	$\mathcal{O}$	$n = 256$
Rayleigh Sommerfeld	n^4	4,294,967,296
AS/F Convolution	$n^2 + 2n^2 \log_2(n^2)$	2162688
Standard Fresnel Diffraction	$2n^2 + n^2 \log_2(n^2)$	1179648
Simple Fraunhofer Diffraction	$n^2 \log_2(n^2)$	1048576

Here, n is assumed to be the size of the side of a square used to represent a 2D image such that the number of pixels would be given by $N = n^2$. The $\mathcal{O}$ notation given here is not strictly accurate as lower order terms have been left in for the purposes of inspection. However, it is clear that the RS method grows much faster than every other method $\mathcal{O}(N^2)$ making even moderately sized experiments untenable even on modern hardware. The Fraunhofer method is the most efficient growing logarithmically with $\mathcal{O}(N\log(N))$, but is not useful as it only gives us the ability to view the far-field diffraction pattern. The convolution methods and the standard Fresnel approach both grow with $\mathcal{O}(N)$ as multiplications between the spectrum and convolution kernel cannot be avoided. This is a significant improvement on the RS method, however.

2.6 Rotational Transforms

As an addition to the above, it is possible to define a transform which has the ability to rotate a field such that you are no longer constrained to propagate along the optical axis or at a normal to the optical axis. This is achieved through modification of the spectrum of a wave, as well as shifting and resampling as in [22].

Applying this in simulation could have great consequences on the efficiency attainable for patterns made of planar segments. Leseberg et al [23] has already looked at creating a surface from non-planar segments with a standard CGH. Were this to be applied with an iterative optimisation algorithm, simple geometries would be easily attainable. Delen [24] and Onural [25] both expand on implementations of this kind of transform.

We have found that these methods cannot be applied in most useful cases. Unfortunately, these transforms have a significant computational burden in order to calculate patterns for high diffraction angles. Furthermore, there is the potentially of loss of information due to the re-interpolation required.

The grating equation can also be used as a metric to understand the required sampling for a sample's wave. As the angle of diffraction is increased, i.e. the wave field is tilted, the required feature pitch also increases. For a field with initial bandwidth/angle of θ rotated by an angle ϕ we may write:

$$\frac{n \sin(\theta + \phi)}{\lambda} = f_{\max} \tag{2.26}$$

By sampling theorum, minimum sample rate is given by

$$\frac{\lambda}{n \sin(\theta + \phi)} = \delta x_{\max} \tag{2.27}$$

For any significant rotation angle (10-40 degrees for example), this number quickly scales beyond use, especially in comparison with the scale of image we are initially considering (i.e. θ of approximately 3 degrees).

Unfortunately we have yet to find a solution to this problem which does not require this increased sample rate to be taken into account within our iterative algorithm, leading to problems which cannot be overcome without vast increases in computing power.

2.6.1 Apertures Transforms

To better model the forward and reverse transforms in the planar case, it is possible to apply the analytical approximation to the spectrum of a square aperture to the

transform kernel(as would be appropriate in the case of pixelated modulators). This better models the diffraction pattern of the modulation device in the case of a Spatial Light Modulator (SLM), such as the ones described and used later in this thesis. Furthermore this curtails some of the noise generated by sampled approximations of the modulator and image. As will be discussed in sections 5.8.1 and 7.3, we have investigated applying a pixelation algorithm to better model and found some useful results.

2.7 Analytical Hologram Design

As has been shown in depth by [26] and [27, 28], applying the FDF analytically leads to some interesting holograms which can be used as a basis for building up more complex patterns. Not only this, but modifications to the propagator have led to analytical solutions to these holograms giving 3D patterns. Some examples are given here as they led to a good basis for initialization for the iterative optimization routines.

2.7.1 Zone Plate

Substitution of $U(\xi, \eta) = \delta(\xi)\delta(\eta)$ where δ is the Dirac delta function into the Fresnel transform leads to the equation for a function that focuses to a point. Explicitly:

$$H = exp\left(\frac{-i\pi(x^2 + y^2)}{\lambda z}\right) \tag{2.28}$$

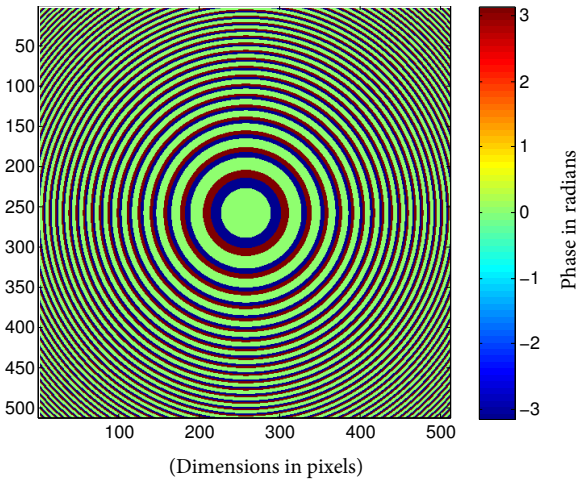


Figure 2.4: Binary phase approximation to a zone plate hologram

2.7.2 Line Holograms

Planar Line

Similarly to the zone plate, substitution of the function $U(\xi, \eta) = \delta(\eta)$ into the Fresnel transform allows generation of a function which can focus to a line on a plane. Limiting the extent of the pattern to a finite distance centred around the line allows a form of length control [27].

$$H = exp\left(\frac{-i\pi(y^2)}{\lambda z}\right) \tag{2.29}$$

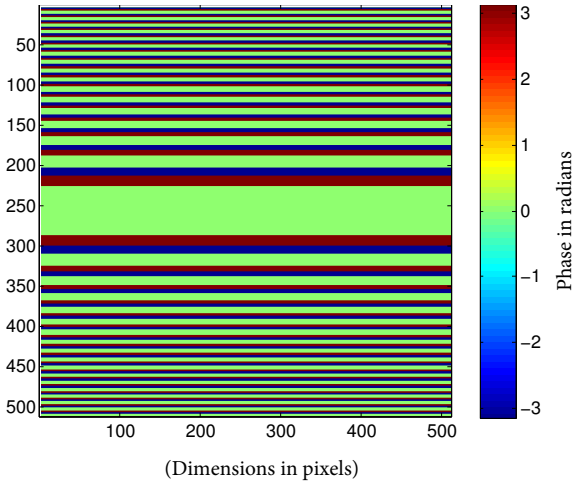


Figure 2.5: Binary phase approximation to a line hologram

2.7.3 Other Analytical Solutions

These equations may also be generalized to lines in 3D, as well as curves as long as a set of constraining limits are taken into account. Maiden [26] also shows how it is possible to introduce width and length control terms to these to gain greater control over the geometry produced. However, this usually requires the inclusion of amplitude modulation terms.

2.7.4 Limitations in Analytical Hologram Design

Used sparsely and at a reasonable distance from the modulator these phase-only patterns can provide good line patterns for lithography. This is discussed in great detail by [26].

Taking a simple analytical line hologram (Figure 2.7) which is $4096\mu m$ square with a focal depth of $8cm$ and simulating its image in a well-sampled regime ($2\mu m$ sample pitch, padded to at least twice the size of the intended image and hologram

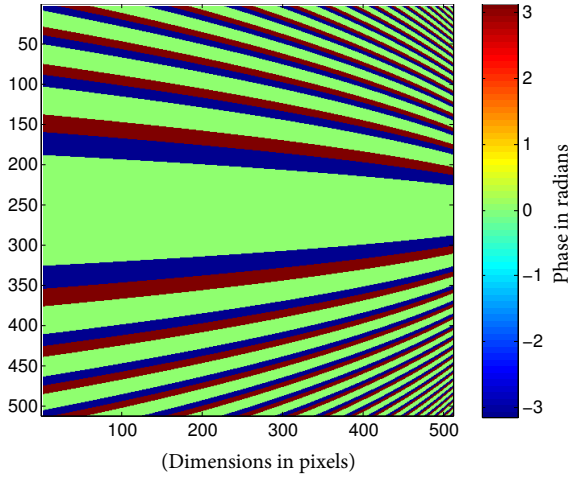


Figure 2.6: Binary phase approximation to a tilted line hologram

size to ensure accurate simulation) we see good contrast and well-defined, even patterns with oscillatory variation most severe the ends of the line pattern (Figure 2.8).

However there are two main problems with this approach. Firstly, addressing the shape of the line patterns in two dimensions. Taking the intensity pattern and simulating a photoresist exposure, we find that the ends of the pattern do not necessarily terminate neatly (Figure 2.9, Figure 2.10) , leading to difficult prospects when considering electrical connections. This can be mitigated to some extent by changing the scale of the image, an image with a longer line length has a proportionally smaller termination region, which is an approach that has been used to some success in [29].

Secondly used in combination, under a similar phase only constraint, the line patterns quickly begin to destroy each other and lead to large intensity variations. When combining multiple phase-only line holograms in superposition (i.e. summed) the pattern goes from phase only with a fixed amplitude, to complex, requiring both modulation of phase and amplitude. If this complex amplitude is achievable in the hologram modulation, the output pattern also becomes a simple superposition of line images. However, if we want to display this pattern of lines with phase-only modulation, we must discard the amplitude pattern. Doing so results once more in a phase-only pattern, but at some degradation of the combined intensity image of the neighbouring lines.

Figure 2.11, shows in simulation how relatively sparse lines perform well, but this soon breaks down as shown in figure 2.12.

Furthermore, the scope for the positions of these line patterns in space are limited. Moving the pattern close to the modulator quickly begins to destroy edge definition and pattern quality. This, therefore, has quite a serious effect on the

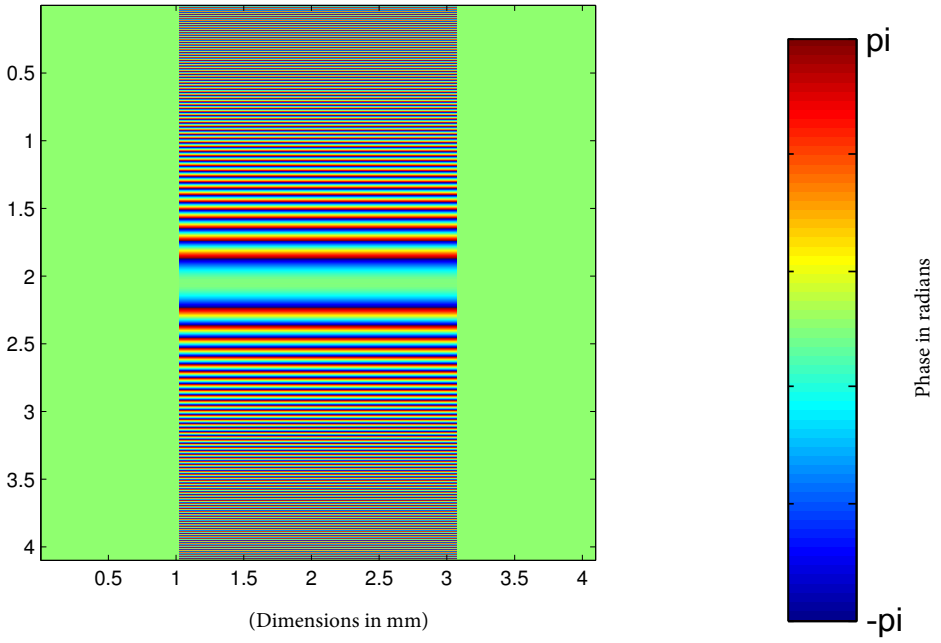


Figure 2.7: A typical un-quantised length limited phase hologram.

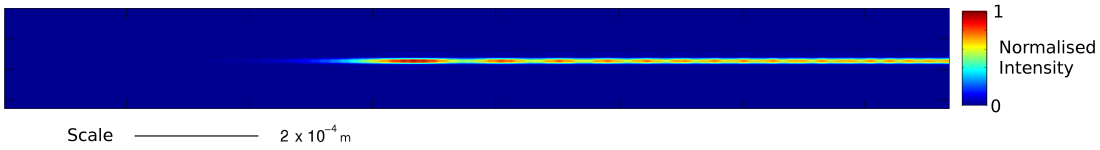


Figure 2.8: A simulated analytical line hologram image

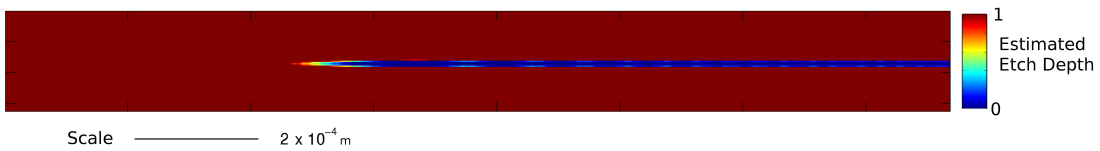


Figure 2.9: A simulated line hologram photoresist profile

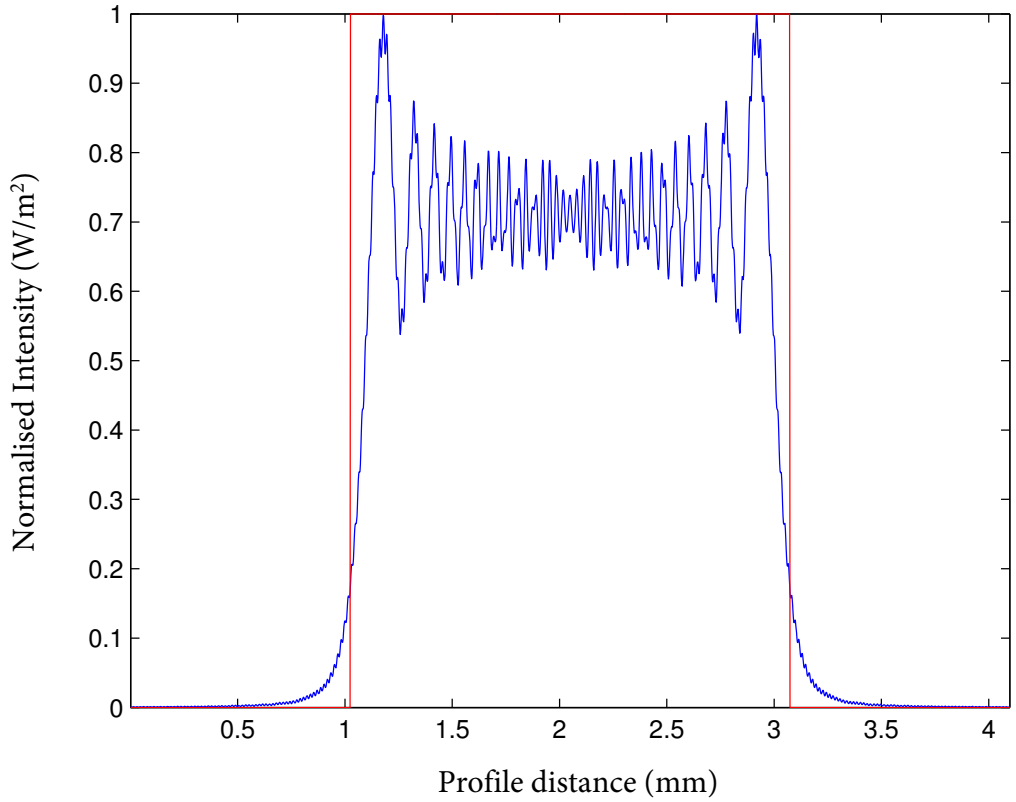


Figure 2.10: Line intensity compared to ideal profile

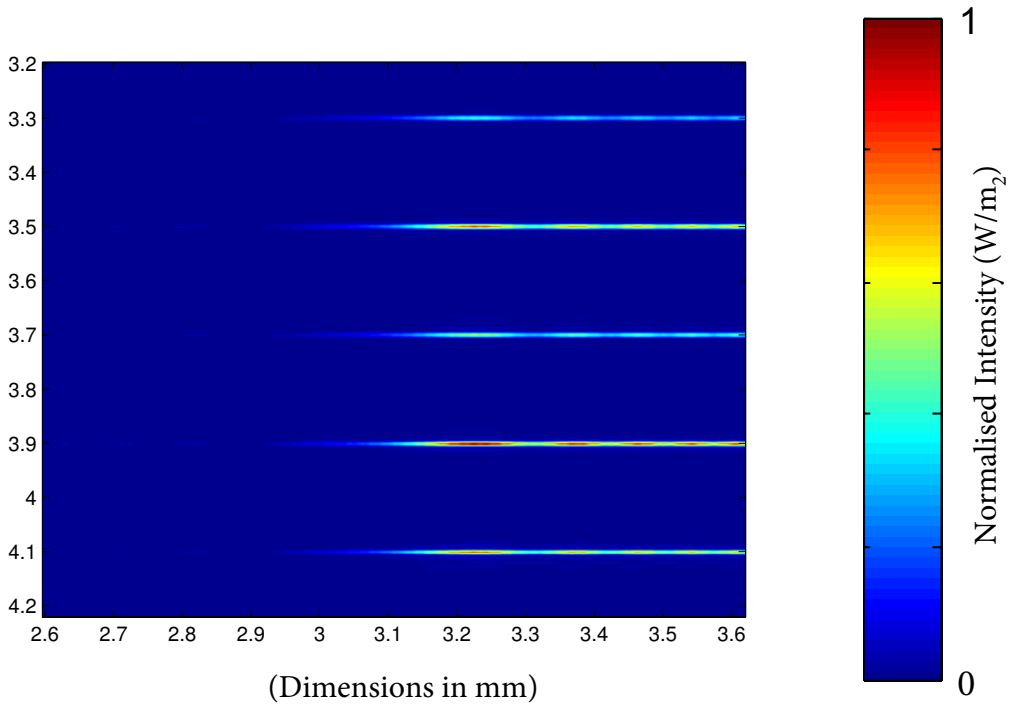


Figure 2.11: A simulated focal intensity pattern for a sparse line bus (from analytical phase only hologram). Line pitch is $200\mu m$. Interaction between line holograms is minimal leading to minor variation between lines.

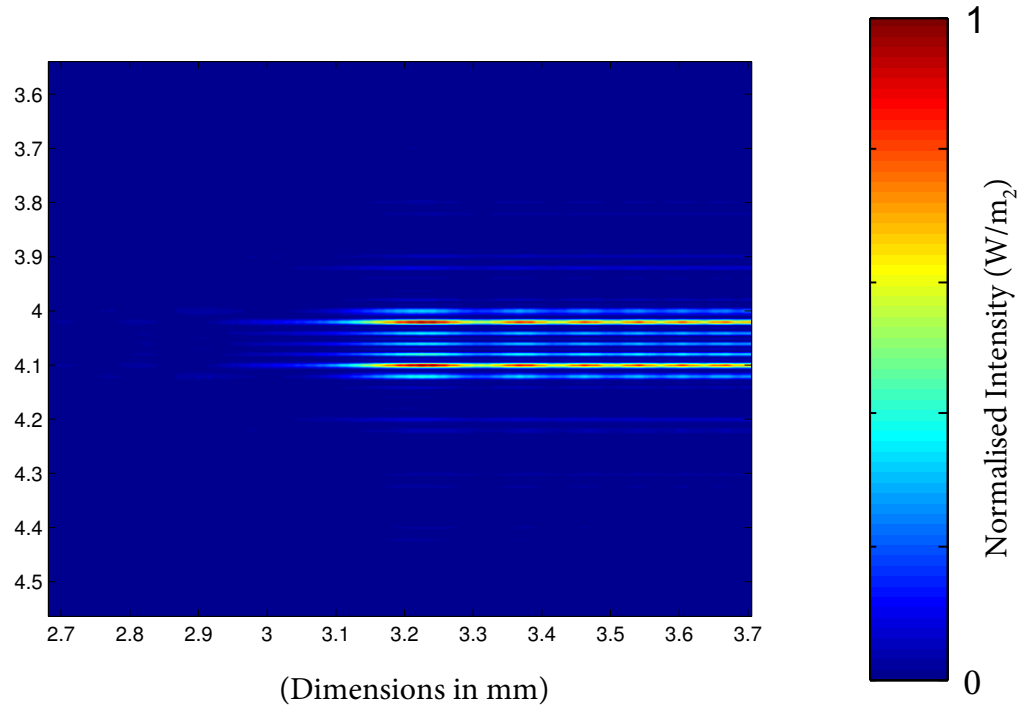


Figure 2.12: A simulated focal intensity pattern for a dense phase only line bus hologram. Line pitch is $20\mu m$. Significant interaction between patterns is unavoidable, leading to significant intensity variations between lines, changes in line profiles and more prominent 'ghosted' line patterns likely due to beating signals between the multiple phase patterns.

types of phase only patterns that can be usefully generated with these simple analytical approximations.

It is important to note here that we have limited ourselves to phase only hologram modulation rather than fully complex modulation. This decision was taken for several reasons. Firstly, electro-optical modulators that have a full range of phase and amplitude modulation do not yet exist. (See [30] for a helpful review of recent investigations into methods to combine amplitude-phase modulation). These active methods often involve multiple modulators or some approximation of complex modulation by grouping multiple pixels. Amplitude and phase passive modulation is expensive and difficult to produce (discussed in more detail in section 4.3). Secondly, including an amplitude term reduces diffraction efficiency, which for the laser sources we wish to use is already a significant issue due to reflections from optical components in the system. Thirdly, the phase-only constraint meshes well with the idea of phase freedom, as applied in many of the iterative optimisation algorithms commonly utilised in optical optimisation (discussed in section 3).

Other modifications - such as the possibility to use multiple modulation devices, or to produce passive modulation in a industrial context which is not sensitive to such high production costs - might lift this requirement, however.

Working under this phase constraint provides valuable information about manufacturable useful design parameters, allows a fast experimental turn around time with modern phase only SLM devices and furthermore opting to optimise towards a phase only pattern will allow us to increase mask diffraction efficiency.

2.8 Conclusions

Two areas have been examined in this review of coherent optical theory. Firstly optical propagation and simulation techniques, from which we have considered their limitations in terms of accuracy and also computational cost. Secondly we have considered the base approximations which allow the creation of analytical phase only holograms.

From examining the propagation transforms, for most simulations considered within this thesis, it has been decided to implement the angular spectrum kernel due to its accuracy and flexibility, as well as being a reasonable computational tradeoff between the RS and Fresnel approximations (see section 2.5.3). However, as discussed in section 5.7, there are some issues with its application for longer propagation distances. Furthermore, in some cases we have found that intentionally limiting the bandwidth of the propagation kernel allows us to abide by practical

constraints through the iterative algorithms we have applied 5.8.1. We have shown how this AS transform kernel can be generated and described some of its limitations examined it computational complexity. We will later show the limits of AS propagation using a local spatial frequency approximation which is expanded upon in section 5.7.

By creating basic phase only hologram constructions such as a line and a bus at 'low' and 'high' densities we have seen how the pattern can break down under a dense pattern and phase only constraint. Despite this, lithographic patterns are still viable and useful exposures can be created. In the next chapter we will discuss hologram optimisation as it exists in optimisation for display (and, in some examples, lithography), and begin to examine how these might be adopted for a 3D lithography technique.

Chapter 3

Review of Computational Search and Optimisation in Holography and Lithography

3.1 Introduction

Analytical solutions to diffraction equations have a place in the creation of sparse line patterns with few intersections. Maiden [26] has shown that analytical solutions can be useful in providing certain patterns of focal lines in space and also investigated many of the parameters of these solutions with their effect on the field and their application to lithography. However very few patterns have neat analytical solutions. The analytical solutions derived through the Fresnel approximation share the advantage that they can be fully represented by phase-only patterns. Combinations of these patterns however result in phase and amplitude modulation being required to form images of sufficient contrast. One advantage of line and point patterns is the symmetry and one dimensional variation of these patterns which leads to good image quality, even under harsh pattern quantisation. The superposition of analytical patterns is a clear path to extending this basic methodology, but bringing image elements into close proximity quickly leads to overlap in their diffraction elements. This in turn brings in requirements for complex modulation or, alternatively, penalties in image uniformity and noise.

To circumvent this limitation, we have several options. The solution space for a hologram with a useful number of pixels quickly expands with size due to the nature of the problem. Even with binary quantisation in mind the full exploration of this space becomes infeasible to compute in its entirety, though this could be considered for very small constructions. For a hologram of n by n pixels with 2

possible phase levels the number of possible combinations is 2^N $(N = n^2)$ and for each one of these patterns a simulation of minimum $\mathcal{O}(Nlog_2(N))$ would need to be computed to find the output image. This ignores any image analysis computation. The total computational cost for analysis of an entire solution space is therefore approximately

$$\mathcal{O}(2^N \times N \log(N))$$

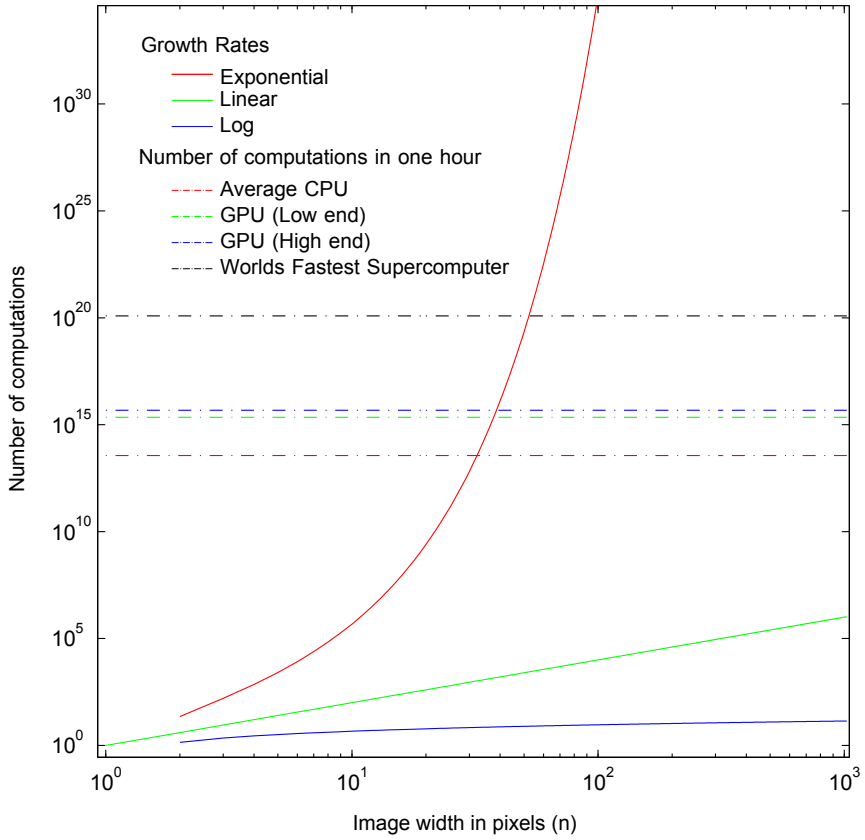


Figure 3.1: Comparison of computational growth with problem size on log-log axis. Vertical axis shows number of computations required, horizontal axis shows size of linear dimension n where $N = n^2$ is the number of pixels or elements in an image. Solid red line shows exponential growth (2^N) as a brute force simulation would require. For comparison we have included linear (green) and logarithmic (blue) growth. Dotted lines show numbers of floating point operations possible for various modern hardware configurations in one hour. Modern CPU (dotted red), modern GPU (dotted green and blue), world's fastest supercomputer (dotted black)

Figure 3.1 shows how the size of this computation grows in comparison to some other common situations, and also how this computation stacks up against modern computing power. The point where two red lines cross indicate the hologram size for which the solution space can be computed within one hour on a modern CPU. This is approximately a linear dimension of 32 pixels, giving a total of 1024 pixels.

This is a best case theoretical maxima, and indeed in practice the computation time is much longer. Moreover, a 32 by 32 pixel hologram is too small to be useful for lithography in practice. Consider a more practical size e.g. $n = 1000$ and this theoretical computation time (for a modern CPU) rises to 4.7×10^{290} *years*. Which is approximately 10^{150} times the expected lifetime of the universe until heat death.

It is therefore clear that this exhaustive approach is computationally unfeasible.

A third approach is to take the form of an inverse propagation or an optimisation algorithm. Simple inverse propagation with no consideration of the pattern phase will often lead to a complex modulation, which when quantised begins to break down, though some work has gone in to finding phase patterns for certain image structures [31]. The simplest optimisation approach might be to one-endedly vary an input distribution and calculate its effect on the output (as in the case of a binary search algorithm). Standard optical optimisation algorithms exploit the phase freedom afforded to a complex modulation pattern to drive the intensity towards a more optimal solution (as in the cases of the input-output algorithm or the Gerchberg-Saxton algorithm).

This chapter presents an overview of these basic approaches in conjunction with discussion on their advantages, disadvantages and quirks.

3.2 Standard Lithographic Optimisation

Before we move to full iterative optical optimisation, it is interesting to note some of the approaches that have been applied in lithography in order to optimise intensity patterns and to push the boundaries of image formation at the smallest scales. These techniques are often categorised under the broad label of "Resolution Enhancement Technologies".

3.2.1 Optical Proximity Correction (OPC)

The method of OPC [6] typically involves adding features (amplitude or phase) to an existing binary mask, such that generally ignored edge diffraction and coherent interference effects behave favourably to the intended pattern. For example squaring off a rounded corner by adding an offset square feature, or adding bars either side of a longer feature to correct feature size and improve definition. Our group at Durham and Sheffield has previously shown that considering these OPC features can be used to begin to compensate for non-planar surfaces [32].

OPC itself can be generalised further by considering "Inverse Lithography" [33], a process whereby an ideal pattern is transformed in reverse. Both process-based

and optical transfer functions result in an ideal mask that hopefully takes these processes into account. However, in these cases the mask is very likely to become much more complex.

3.2.2 Source-Mask Optimisation

Modern lithography processes may exploit their ability to change the shape of the source pattern to adjust the coherence of the beam [33]. This can lead to reshaping of the point spread function of the system which in turn affects the geometry of the pattern. Often, as with OPC, this method is used to turn an otherwise filtered and rounded Manhattan geometry interconnection pattern back into a sharp rectangular image.

3.3 Review of Iteration

Certain algorithmic approaches and their derivatives provide an interesting technique for hologram design. In opposition to the rigorously analytical solutions which are then approximated by band limit and quantisation, these iterative methods often start with a random seed and use "error-reduction" algorithms to search the possible solution space for a valid hologram.

Because of the nature of this method, the solution converged upon is unlikely to be a globally optimal solution. Certain modifications, however, do account for this and attempt to improve the convergence of the algorithm by avoidance of stagnation into local minima [34].

These algorithms can have important effects on the holograms generated as they can approach a solution whilst maintaining many separate limiting conditions which allow better modelling of a modulation device, such as in the case of quantisation (see section 3.4). They also give the ability to produce patterns which have no known rigorous analytical solution, and are especially helpful in cases where a global optimum is not required where an approximate solution is still useful.

Many of the optimisation processes used in the optical optimisation algorithms discussed in this chapter are based around the idea of "phase freedom". This means that the intensity of the pattern is the part of interest which does not depend on the phase pattern in the image plane. As such this phase is applied as a free variable allowing any underlying phase pattern as a valid solution and so a particular image can be generated by a large number of possible complex input patterns. The core of these algorithms is the Gerchberg-Saxton algorithm [35] which we will discuss shortly.

3.3.1 Error metrics

Before considering the iterative optimisation techniques, it is important to consider error and quality metrics.

Many different figures of merit may be applied to the holograms generated either analytically or numerically. Practical lithographic measures might be line width regularity and defect rate per area. As the project moves forwards, away from research and towards engineering development, it is expected that these measures will become more important.

However, at the current stages of development, signal processing and optical measures are appropriate for analysis of the different image patterns generated by our methods. To this end some of the more important measures are noise quantification by standard deviation or measures of variance, as well as 'contrast,' which we will discuss later in this report.

The standard error metric for image comparison within the context of the algorithms presented here is that of the sum of squared error (SSE). Since, for synthesis, the iteration methods described below use and ideal image which is optimised towards this creates a useful basis for comparison.Fienup [36] cites this basic normalised error measure.

$$E = \sum_{x=1}^{N} \frac{|g(x) - f(x)|^2}{|f(x)|^2} \tag{3.1}$$

Where E (the error) is determined from a sampled function corresponding to the ideal pattern $f(x)$ and a similarly sampled function for the questionable image $g(x)$. The work of Fienup goes on to describe how to eliminate irrelevant error for certain hologram implementations, such as twin image, zero order noise, translation and multiplicative factors. These issues are largely irrelevant for the phase only holograms we consider in this chapter.

3.3.2 Gerchberg-Saxton

The Gerchberg-Saxton algorithm [35] is the basis for much of the development for many more of the iterative algorithms presented here. Its initial invention was intended to regain phase information from multiple intensity measurements within the context of electron microscopy. The iterative loop of the algorithm is as follows, and a flow diagram outlining the algorithm can be seen in figure 3.2.

- Starting with two sampled intensity measurements (often image and diffraction planes), $I_A(x, y)$ and $I_B(\xi, \eta)$, we define the measured amplitudes as $A(x, y)$

and $B(\xi, \eta)$, where $A = \sqrt{I_A}$ and $B = \sqrt{I_B}$. Where (x,y) are coordinates in one plane and (ξ, η) are coordinates in a corresponding Fourier plane.

- Begin by generating a random phase pattern $e^{j\phi(x,y)}$ as a guess at the initial phase of our complex function.
- Multiplying the amplitude function $A(x, y)$ by the phase function $e^{j\phi(x,y)}$ as a guess at the fully complex wave. $A(x, y)e^{j\phi(x,y)}$ i.e. impose the amplitude constraint on the initial plane.
- Taking the Fourier transform of this function we arrive at a complex function now in the coordinates (ξ, η). $\mathcal{F}(A(x, y)e^{j\phi(x,y)}) = A'(\xi, \eta)e^{j\phi(\xi,\eta)}$ (Note: this can be represented in Cartesian coordinates, or in terms of diffraction angle)
- Now we replace the propagated amplitudes with the ideal second amplitude pattern giving $B(\xi, \eta)e^{j\phi(\xi,\eta)}$.
- Compute the inverse Fourier transform, moving back to the initial coordinates. $\mathcal{F}^{-1}(B(\xi, \eta)e^{j\phi(\xi,\eta)}) = B'(x, y)e^{j\phi(x,y)}$. The process is then repeated from application of the constraint to the A plane.

The GS algorithm is intended to recover a phase pattern from two intensity measurements. In this case, these intensity measurements are that of a diffractive element and its corresponding far field diffraction pattern. For optical cases this can be achieved by looking at an image created at the focal plane of a lens.

The core functionality of the algorithm is generated by the proof in the original Gerchberg and Saxton paper showing that applying the algorithm can *never* cause the sum of squared error in the image to increase.

This algorithm can be modified to the case of phase hologram synthesis [37–40], and it is especially useful for modern phase only optical modulators. To do this we enforce that the diffraction plane must have unit amplitude for all (x,y) samples, leaving the phase as the only free variable to modulate the field.

It must be noted that the basic Gerchberg-Saxton algorithm does not necessarily optimise holograms with regards to any measure other than the total SSE. The original paper demonstrates how the algorithm can only decrease the SSE for a particular image pattern. This optimisation does not however guarantee that iterated solutions will not develop local defects or patterns that have a low effect on the SSE whilst improving this measure on the image as a whole which is a topic we investigate later. We have found these algorithms generally results in a noisy intensity pattern for anything other than 'thin' features (and even then only under certain initial conditions).

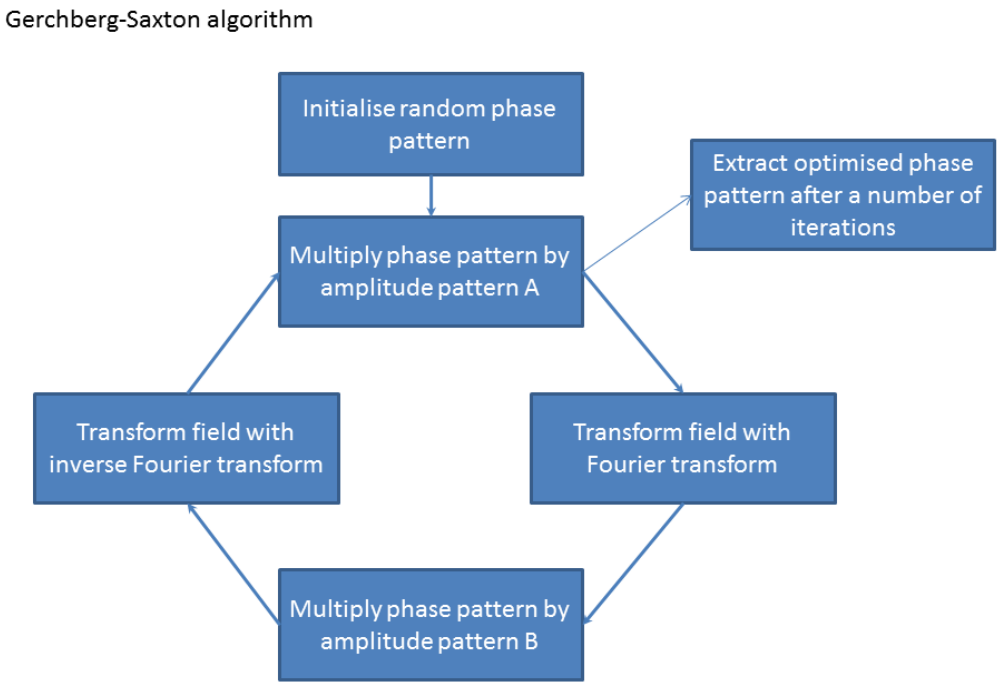


Figure 3.2: Flow diagram outlining the key steps in a basic Gerchberg-Saxton Algorithm

Despite this 'non-optimal' optimisation, as discussed in chapter 5, the GS method and its derived algorithms are very good at containing energy within a given region. In conjunction with some other simple modifications, it can indeed give viable results for lithography.

In the case of implementations of all of the following error reduction techniques a convergence criterion must be chosen to decide upon when to stop the iterative improvement. For all of the implementations of the GS algorithm and its derivatives which we have explored in this thesis improvement is highly non-linear with number of iterations. In most cases 10-20 iterations will give the bulk of the improvement. Though the algorithm does not usually completely stagnate improvement beyond this level it is slowed significantly. Therefore for most experiments in this thesis we have chosen a fixed number of iterations 50 to 100 which have been sufficient to observe the bulk of the improvement in a reasonable time frame.

3.3.3 Fienup GS variants

Steepest descent

In the "steepest descent" method [41] a fundamental approach to numerical hologram design is considered. In this method each pixel is treated as a separate variable and we may calculate partial derivatives for each of these inputs with respect to an image quality metric. We can then find the vector along which the "steepest gradient" exists and move the solution accordingly.

Fienup states in his comparison that the GS algorithm is actually a special case of this steepest descent algorithm which utilizes a fixed step size. In this case a significant advantage is born of the applicability of the FFT algorithm which can be calculated with greater efficiency.

Input-Output

The Input-Output algorithms described in [41,42] borrow well-known methods from the field of astronomy as error correction algorithms. The same core methodology of Fourier transforming back and forth between two distinct domains and imposing constraints on each takes place. However, the error recorded in the image domain is corrected directly. Other constraints are also applied to the image or its Fourier counterpart for example, the field values must be positive and may be confined to a set region. Quoting from [41]:

$$g_{k+1}(x) = \left\{ \begin{array}{cc} g'_k(x), & x \notin \gamma \\ 0, & x \in \gamma \end{array} \right\} \tag{3.2}$$

$$g_{k+1}(x) = \left\{ \begin{array}{cc} g_k(x), & x \notin \gamma \\ g_k(x) - \beta g'_k(x), & x \in \gamma \end{array} \right\} \tag{3.3}$$

The above equations show two different possible methods for calculating a next image estimate $g_{k+1}(x)$ from previous estimate $g_k(x)$, and new estimate $g'_k(x)$ which has been modified by Fourier domain constraints.γ is the set containing all points where the constraints are violated. β is an arbitrary constant chosen to influence the convergence characteristic of the algorithm.

Investigating these "error reduction" and "input-output" approaches, Fienup discusses the interesting result that any combination of the above choices for areas in γ do give convergent behaviour to some extent. He noted that one particularly useful combination is that of

$$g_{k+1}(x) = \left\{ \begin{array}{cc} g'_k(x), & x \notin \gamma \\ g_k(x) - \beta g'_k(x), & x \in \gamma \end{array} \right\} \tag{3.4}$$

i.e. an amalgamation of the above methods. But of particular interest is his conclusion that by changing the method used every few iterations the convergence improved beyond what any individual method could achieve, by enabling the algorithm to avoid stagnation.

3.3.4 Iterative Fresnel Algorithm

Taking the GS algorithm further, Dorche et al [37] have shown that it is possible to achieve convergence without need for the explicit Fourier transform such that the same effect can be exploited when applying the Fresnel transform (which does incorporate a Fourier transform). This allows much more precise control over the depth of projection of an image. Dorche also explicitly investigated this algorithm as a tool for image synthesis rather than phase recovery. His experimental reconstructions of images at the focal plane of a lens and propagated away from it show some of the first examples of this algorithm being used for 3D modulation of a pattern.

Utilising the Fresnel transform, and phase only patterns also allows us to create patterns with no (collected) zero order noise and high diffraction efficiency. Using lens-less propagation also allows holograms generated using such a method to be incorporated into existing proximity mask aligner systems with the minimum of modification, since the image is then not in the far field or at the opposing focal point of a lens.

This method also allows comparison of this method directly against the similarly lensless analytical structures generated by techniques such as the cylindrical patterns in [27, 28].

Fresnel 3D Derivatives

Makowski and Sinclair [38,43] both make use of the iterative Fresnel scheme to begin to generate 3D patterns for specific applications. Dorche [37] was the first to suggest an iterative algorithm based around free space propagation routines. Beam-shaping elements and holographic optical tweezers are cited as possible applications for this 3D iteration, though some interest has been shown in 3D display [44, 45]. Makowski et al [38] specifically evaluates the numerical propagation within the iteration with the convolution form of the Fresnel integral. To supply 3D control over the wave field, the G-S algorithm is generalised to include multiple output planes just as

Dorche did. The images produced by this system seem to be of good quality for a reasonable separation of output planes.

Effective "crosstalk" between the images occurs when planes are brought into close proximity. This may occur due to constraints on the output volume which break the constraints of the Ewald equation on a 3D monochromatic field. Because of this, a low-error solution cannot be found under the terms of the scalar wave equation. For example, applying patterns on two planes in an image volume close together may require higher frequency information than can be provided by a wave field at a particular wavelength, or more likely cannot be supported by the bandwidth of the aperture-limited system. Xia [40] explores the use of Makowski's algorithm for 3D display, and cites an interesting technique which we have subsequently adopted and adapted (See section 6.2).

3.4 Iterative quantisation

The constraints imposed by many of the iterative techniques discussed so far are that of phase only patterns in the diffraction plane and a specific amplitude pattern in the image plane. This can be extended more usefully within the context of the iterative procedure. Wyrowski [46] develops - and [47–49] explore the implications of - a quantisation technique which allows the constraint to be gradually applied to the image and hologram, much in the same vein as a simulated annealing technique. This potentially results in a much simpler diffractive element whilst attempting to retain the ability to produce a high quality image. This can help in the physical manufacturing of holographic masks, as the fewer phase/amplitude levels required for projection the easier and cheaper they are to produce.

To do this, they generate an operator which modifies the clipping constraints of the hologram pixels over successive iterations.

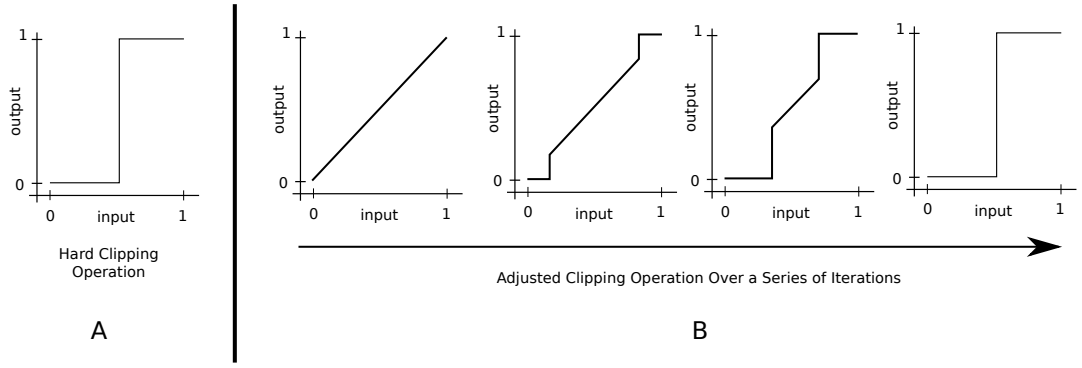


Figure 3.3: A) a so-called "Hard Clipping" operator B) an operator which changes its function over a series of iterations

This clipping operator can be generalised to multiple phase levels easily as in [50]. Skeren et al [50] is able to generalise this approach to a series of four separate approaches to iterative quantisation and shows their comparative merits.

Their conclusion in the comparison of these methods is to say that direct "hard-clipped" methods seem to give as good if not better S/N ratios in output masks than the other more complex iterative quantisation experiments. There is scope for further research into the type of noise generated by the different operators, however, as the more complex methods may well be more suitable for lithography or other purposes.

3.5 Noise in iterative holograms

Certain types of noise are a well-known issue in phase-only hologram optimisation owing to issues in the convergence of the algorithm for phase only holograms. These convergence issues cause optimised holograms to contain numerous nulls which force the intensity pattern to a zero value.

Some [51] have suggested that the source of this issue is the inability of the algorithm to remove 'optical vortices' from a phase field pattern. Following this an iterative process can be defined which includes a vortex removal step once the iterative process has stagnated. Alternatively, Wyrowski [52] suggests that a bandlimited IFTA under a complex set of smoothed initial conditions can significantly reduce the effect of this noise. These methods, however, are complex and computationally taxing, requiring a supersampled field to be processed in order to identify the locations of vortices in the field before applying a phase function which attempts to remove the vortices.

Some, such as [53, 54], instead mitigate this actively by signal averaging. This method is simple but potentially computationally taxing as it requires many separate iteratively optimised holograms to be computed. Furthermore it requires that the modulator used be an active one so that each of the separately optimised holograms can be displayed quickly in sequence, resulting in a desirable average field. Golan and Shoham [55] have demonstrated that in certain situations the requirement on extra computation can be minimised.

Another approach is to afford the iterative algorithm some amount of amplitude freedom in the image plane [56, 57]. Amplitude freedom methods, however, are not particularly appropriate for lithography as they result in significant portions of energy being diffracted into potentially sensitive parts of the pattern.

A final approach is to modify the coherence properties of an optical system to

mitigate the noise by effectively 'blurring' the output image. This can be done by making a coherent source less coherent by the use of rotating diffusers [58] or it can be achieved by increasing the coherence to allow 3D modulation from an effectively incoherent source.

As we will show later, it is possible to iteratively optimise certain patterns without introducing such nulls (see Chapter 5) if pattern constraints are abided. Additionally, the results of coherence and signal averaging techniques simulated or tested with our 2D and 3D iterative algorithm will be presented later in this report.

3.6 Direct Binary Search

The direct binary search (DBS) algorithm [59, 60] is not related to the other search algorithms described here, and essentially closely resembles brute force search for a binary pattern evaluated pixel by pixel. Each point is taken in turn and flipped from one value to another. This could be transparent to opaque or $0,\pi$ phase. The change in output error is then calculated and the pixel value with the lowest error is retained. This can continue until the error stops reducing or within a certain error range. This algorithm has the potential to find an optimised solution under serious binary constraint which is desirable but is far from computationally efficient, though some (such as [60]) do attempt to speed up the process.

3.7 Utilising the Ewald's Sphere

Whyte [19], considered another generalisation of the GS algorithm in which a whole 3D field is represented and constrained simultaneously. Physical and frequency space representations of a volume are sampled and the constraints of the image volume and Ewald's equation are applied. Once the image volume has been iterated for some time, the diffractive plane is generated by a transform from the Ewald's sphere's surface. This method has the advantage that 3D fields are considered from the outset and it allows the formulation of complex 3D objects. From the results presented in [19] however, the volume intensity pattern formed are vague suggestions at the shape rather than explicit 3D objects. Also, a vast memory requirement in evaluation of a sampled 3D volume limits the practicality of this method.

3.8 Optimised Photolithography

In current cutting edge photolithography several approaches can be made to utilising wave theory in lithography. Resolution Enhancement Techniques (RETs) [6] and inverse lithography (ILs) are two such approaches. As discussed previously 3.2.1 these two are similar in that both try to start to take advantage of the diffractive properties of light to improve resolution/definition in an exposure. Certain RET techniques show how serifs can be applied to masks and simply add fringes to the pattern to give a more defined output. In general, IL utilises a reverse propagation from substrate to mask. This can be used to generate features similar to serif patterns but more general in scope.

Though these RETs utilise diffraction to improve the resolving power of standard shadow-mask setups, only a few have attempted to apply iterative diffractive optimisation to of holographic lithography specifically. The most prolific publisher in this area seems to be Frank Wyrowski [61], whose approach is to take a partially coherent model and a method called "Projection On To Convex Sets"; another iterative approximation algorithm.

Recently another paper has emerged approaching the topic of 2D holographic lithography as well also from an iterative algorithm approach and using the GS algorithm [54]. Their approach is to generate many partially optimised holograms and to counter the noise by cycling the mask displayed on an SLM over the course of the exposure. Whilst this may work well for use on an SLM, in real situations, the resolution may need to be higher and the exposure area larger, thus restricting the use of an active modulation device. This approach has been studied previously by [53]. We also have applied it to suppress noise using our SLM as seen in section 5.12.2.

Another such optimisation algorithm is that of the Intel engineers in [62]. This optimisation procedure operates with a binary mask and utilises pixel sizes smaller than a single wavelength. This allows calculation of patterns with a very attractive profile for lithography. Also, this method introduces several other interesting techniques such as stitching together of neighbouring optimised holograms, and optimising out undesirable mask features such as single surrounded inverted pixels, which can cause mask sensitivity to damage.

All of these lithographic techniques, however, are only interested in the planar case, and not in lithography in 3D, which is the topic we intend to approach from here onward.

3.9 Conclusions

In this chapter we have discussed a range of iterative optimisation algorithms which can be applied to the design of holograms. Many of these have been derivatives of the Gerchberg-Saxton algorithm, which we will go on to practically assess in a simple test case in Section 5.3. We have also seen that algorithms exist for manipulation of a 3D optical field with a phase hologram, but the limitations of these algorithms are yet unclear. The manipulation of such algorithms into a form which can be used for lithography will form a central tenet of the rest of this thesis. Methods of constraint to phase only or quantised phase patterns have been used extensively for display. Iterative optimisation of phase patterns has also been adopted in techniques such as 'inverse' lithography for binary structures at the cutting edge of lithography technology. This suggests that more extensive manipulation of the phase of the optical field is potentially not too large a divergence from existing methodologies to be useful. To our knowledge, none have applied any of these optimisation techniques common to 2D lithography to 3D patterns in lithography, and none investigating 3D optimisation of light fields in areas such as display have attempted to apply these methods to lithography. This is the gap that we wish to investigate, i.e. taking the optimisation of 3D field by a Gerchberg-Saxton-derived algorithm and applying it to create useful non-planar lithographic structures.

Chapter 4

Optical setup for holographic lithography

4.1 Introduction

We have now outlined the theory of optical propagation and methods to generate iteratively optimised phase only holograms. This chapter goes on to examine the implementation issues involved in setting up a practical holographic modulation for lithography. Furthermore, it discusses how this affects the design of the system and optimisation, and makes some conclusions which limit the scope of this project as a consequence.

4.2 System Overview

Traditional holograms, computer-generated or otherwise, have a variety of encoding mechanisms and optical regimes which are applied for different purposes. On- or off-axis reference beam, Fourier or free space, phase or amplitude modulation etc.

The approach in which we apply holograms is not holography or digital holography in the classical sense of the term. Holography is a way to record and encode phase information in intensity-sensitive media. For a traditional emulsion hologram, the monochromatic beam reflected from an object should be combined by interference with a reference beam (usually a delayed beam from the same source to ensure coherence) causing an interference pattern. This pattern then contains a modulated intensity which is a function of the superposition of the intensities of the two waves and also contains terms which are generated by the complex interaction of the two waves.

In modern digital holography it is possible to use a variety of methods to separate

these components with relative ease [63]. This allows us to reconstruct the object phase field. One may also "re-play" the hologram by illuminating it with a conjugate reference wave allowing observation of a slightly obscured 3D object through the hologram.

Since we wish to project a real 3D object with as little distortion as possible, and we are not interested in recording or displaying an existing 3D scene, we concern ourselves primarily with the projection aspect of holography, i.e. modulating a "reference" wave to project a 'real' image on a substrate. The other aspect of the system - the creation of the hologram for projection - is covered by the iterative and analytical methods described previously. This topic is often termed 'Computer-Holography'.

An optical system capable of such holographic projection for lithographic purposes consists of light source, source conditioning optics, modulator, projection optics, alignment mechanics and substrate, each of which are discussed here in turn.

4.3 Optical Modulation

Modulation technologies can be broadly categorised into 2 groups, passive and active.

4.3.1 Passive Modulation

Passive modulation typically involves a transparent mask substrate which by way of deposited inks, metals or polymers, or by etched geometry, modulates the phase or amplitude or both of an advancing wave. In its most basic form this could be something like a chrome on glass mask which blocks particular sections of an advancing wave front in much the same way as a traditional mask. Using this approach in conjunction with understanding the coherence properties of the wave we are able to form 3D images. These binary amplitude patterns are the simplest from a fabrication point of view, and can work surprisingly well for line and point patterns (i.e. patterns with a phase only analytical solution) as demonstrated in section 2.7. In general, more complex hologram constructions are preferred: The more the modulation pattern is constrained and restricted the more noise is introduced to the system. In practical situations it is more difficult to manufacture anything more complex than binary amplitude masks since there is very little call for accurately produced multi-level amplitude and phase modulation masks outside of research, or cutting-edge nanoscale development.

This can be circumvented to some extent by considering sub-aperture holograms,

such as Lohmann [64] (detour phase) holograms. However to achieve this sub-aperture phase hologram a printable resolution much higher than the intended hologram 'pixel' size is required to achieve the shaped/positioned sub-aperture structures. Moreover, noise terms are introduced by this representation which can degrade the image [65],though work is being done to curtail these issues [66].

At the very cutting-edge level of industrial lithography this low-complexity constraint is lifted to some extent. Multi-level masks are produced more frequently for phase-shift lithography (a form of RET), which uses binary amplitude and binary phase masks, which can improve feature size [62].

4.3.2 Active Modulation

Active modulation devices are a relatively new concept and can be formed from several different technologies. Primarily, these can be categorised as; deformable wavefront modulators, such as those used in astronomical adaptive optics [67]; binary digital micromirror devices (DMD) [68], such as those used in modern digital projectors; liquid crystal on silicon (LCOS) [69] modulators akin to those used in modern liquid crystal displays. All of these approaches are reconfigurable to some level and have different constraints on the kinds of modulation they can achieve.

DMD's are the most restricting because of their pseudo-mechanical construction, being controlled by mirrors which flip back and forth between two positions providing an 'on' and 'off' for each pixel. This leads to a binary amplitude modulation. In modern projectors this binary state can be modulated at a high frequency to provide an analogue average power to the human eye but in coherent optical systems the usefulness of this approach is limited. Deformable mirrors provide analogue phase modulation with a relatively low resolution and tend to be expensive MEMS components.

LCOS devices are capable of pure phase, pure amplitude or some limited set of phase and amplitude modulation states [70].

For further reading on applying active modulation, the affects of quantisation, and a basic overview of active modulators, please see these recent reviews [71–73].

4.3.3 Quantised Modulation

To compare and contrast the methods described above, we need to know a little about how different modulation constraints affect potential image quality. This is not necessarily as simple a topic as it may seem. In general, the more modulation states and the wider modulation range in both phase and amplitude that can be

applied when modelling a field, the better quality the final image. However, some patterns (such as the simple analytical solutions shown in section 2.7) change very little under binary amplitude or phase only constraints. The topic of modulation and quantisation schemes has enormous scope, however. Many have applied phase only modulation as a useful constraint in holography. The importance of the idea of phase only modulation in the context of lithography has several consequences.

Combined phase and amplitude active modulation does not exist, though some have approached methods to begin to overcome this limitation [70,74–76]. This leads to an implicit technological limitation if we require a fast turnaround for experimental results. When testing laser based holographic modulation we have discovered that a significant problem with the optical system is the loss of energy through each optical element in the system. For even straightforward configurations it is possible to lose upwards of 95% of the beam energy through the filter, collimation, and splitter sections of the optical system. With simple amplitude modulation this loss is further amplified. Therefore, an opportunity to minimise further energy loss (by considering phase only modulation as our practical constraint) is of benefit in the short term, though it is clear that such a limitation is an addressable engineering challenge for a future "production" quality system.

Since it is the aim of this thesis to lay a foundation for optical design in real practical holo-lithographic optics another reason for phase modulation presents itself. Though active modulation is in most cases strictly restricted to phase OR amplitude modulation, passive modulators are not. However, complex passive modulation is very difficult to produce or even have produced commercially, as very few applications have a need for such modulation. Therefore, by taking the phase only quantisation constraint we also limit the cost and complexity of production of such passive modulators in future research derived from this work. This will become an important constraint as this technology progresses towards industrial application.

It is for these reasons of cost, complexity, and power that we will consider phase only modulation as our primary constraint in this thesis.

Removal of the amplitude term results in issues removing certain coherent noise problems from the image as discussed, by Section 5.10. We have subsequently overcome this by temporal signal averaging or coherence reduction, as demonstrated in Section 5.12.

4.4 Optical Sources

Traditional mask alignment lithography machines use sources such as mercury arc lamps. Previous work at Durham University laser systems has used equipment such as the Cube 405nm diode laser from Coherent inc. This ensures that the source is sufficiently coherent. Coherence is required to form our 3D patterns which rely on coherent interaction between modulated regions of a hologram. The loss of coherence in such an optical system results results initially in a loss of resolution eventually leading to a largely incoherent superposition of patterns in which no interference effects are observed.

4.4.1 Source Conditioning

Collimation of the source beam makes sense for most hologram projections. This provides the simplest condition for hologram constraint in a simulated iteration (a plane wave illumination) and an image centred about the optical axis. It is potentially advantageous to consider more complex illuminations functions such as converging, diverging or linear phase functions to allow scaling or shifting of the image and to extend the effective bandwidth of the modulation device. However, in order to keep things relatively straightforward and easy to align and construct this has not been considered here. The standard method to clean and project a single simple plane wave illumination therefore is applied, which is to use a spatial filter which consists of a short focal length microscope objective focussed on a pinhole to pass only a very small section of the spectrum of the wave. The diverging pattern from the pinhole, which becomes close to a Gaussian amplitude pattern when higher frequencies are removed. This is then collimated by a larger aperture lens and the central part of the Gaussian illuminates a hologram or other optical elements. This central section of the Gaussian pattern ideally has an almost flat amplitude and phase profile, as would be provided by an ideal plane wave illumination. This is displayed diagrammatically in figure 4.1.

It is interesting to consider how this basic holographic setup differs from a standard lithographic projection system. In a lithographic system, the sizes of the components can become large owing to the need to expose a larger area (as compared to the optics required when using an active SLM modulator). Low resolution masks are also an option which may then be photo-reduced for patterning onto a die. A similar process is used however to condition a source, with a few further additions. A traditional mask aligner might use an arc lamp source, which is likely filtered to a single or set of spectral peaks. This partially coherent beam is then collimated

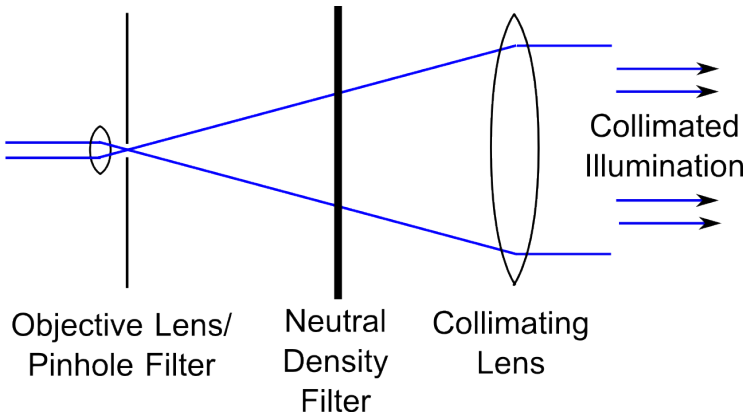


Figure 4.1: Laser source collimation

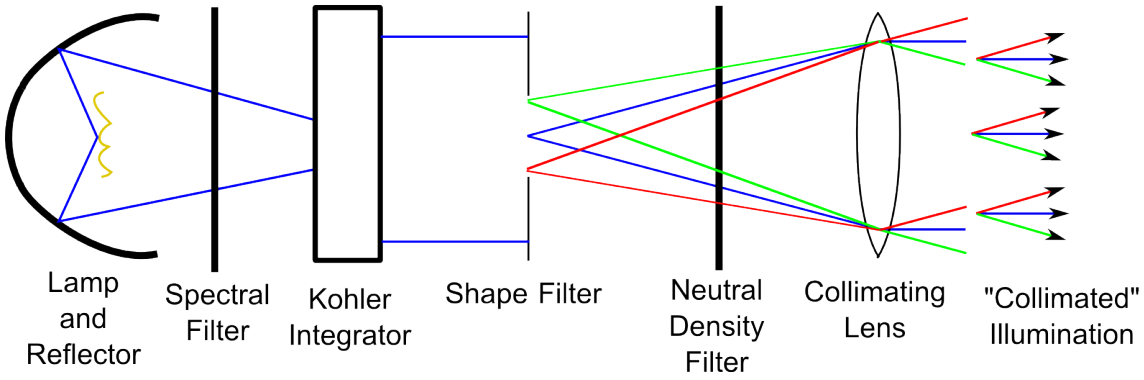


Figure 4.2: Arc lamp source shaping and collimation

and homogenised by a lens system and potentially a set of micro-lens arrays such as in [77]. This will then be projected through an aperture to give a shaped source (giving a partially spatially coherent beam) which is then re-collimated and used to illuminate a mask. The image of the mask will then be projected through a photo-reduction system onto a substrate at the desired resolution. Please refer to [58] for full details on this complex illumination system, but a diagram of simplified shaped illumination can be observed in 4.2.

The important changes are that of the control of the coherence of the system through both spectral filtering and source shaping. These can affect resolution and accuracy on some patterns. It would be possible to emulate the coherence of laser with an arc-lamp source if the spectral bandwidth were limited to a very narrow range, and the source shape filter size was made very small.

Partially coherent systems (as opposed to near or fully incoherent systems) are often exploited in lithography to achieve better feature resolution by phase shifting sections of masks. Some interesting work has been carried out by Volger et al from Suss Microoptic [78] wherein coherent diffractive structures (Fresnel zone plates)

were applied with complex partially-coherent illuminations to form focal patterns over a proximity gap, thus increasing feature resolution.

The bulk of this work does not concern itself with partial coherent or incoherent systems, focusing on what can be achieved with a single collimated laser source, though we are clear that this is a future research topic and is very important to optical systems of this type.

We will, however, later discuss strategies for removing coherent or speckle-like noise from our optical system. This largely involves reducing the coherence of a laser source, OR increasing the coherence of an arc-lamp system (see section 5.12).

4.5 Projection Systems

For our particular application there is one fundamental choice to make with regards the optical projection system past the modulation device. This is the option to include or not include a lens or system of lenses after the modulation device, this decision somewhat equates to the choice between a contact or projection system in a standard lithographic optical system. Figure 4.3 shows diagrams of lensed and lens-less holographic projection.

The inclusion of an optical element in a 2D holographic system is usually applied to project a Fourier hologram instead of a free space hologram at the focal plane of the lens. This however is only the most simple case. By depth propagation away from the focal plane of the lens it is just as possible to create images around the lens focal plane and begin to construct 3D images. We do this by considering one of the optical propagation equations described in chapter 2 in place of a simple Fourier transform as might be used in a far-field case. The lens or lens system post holographic modulation acts as a further phase modulation. For a single image point this is likely to shift the focal plane, as would be expected. This type of system is of potential interest where the hologram bandwidth is insufficient to produce the modulation required for a particular pattern and the hologram contains a strong term similar to the lens phase element.

This lens modulation term can then be dropped from the hologram and result in a bandwidth pattern to be implemented in modulation. However as with any extra optical element, alignment and further aberration and limitation begin to play a role. Depending upon the size of the lens and where it is placed relative to the hologram it may further limit the resolution due to the size of its input aperture, and for very strong lenses aberration terms can disrupt the phase pattern significantly unless corrected for.

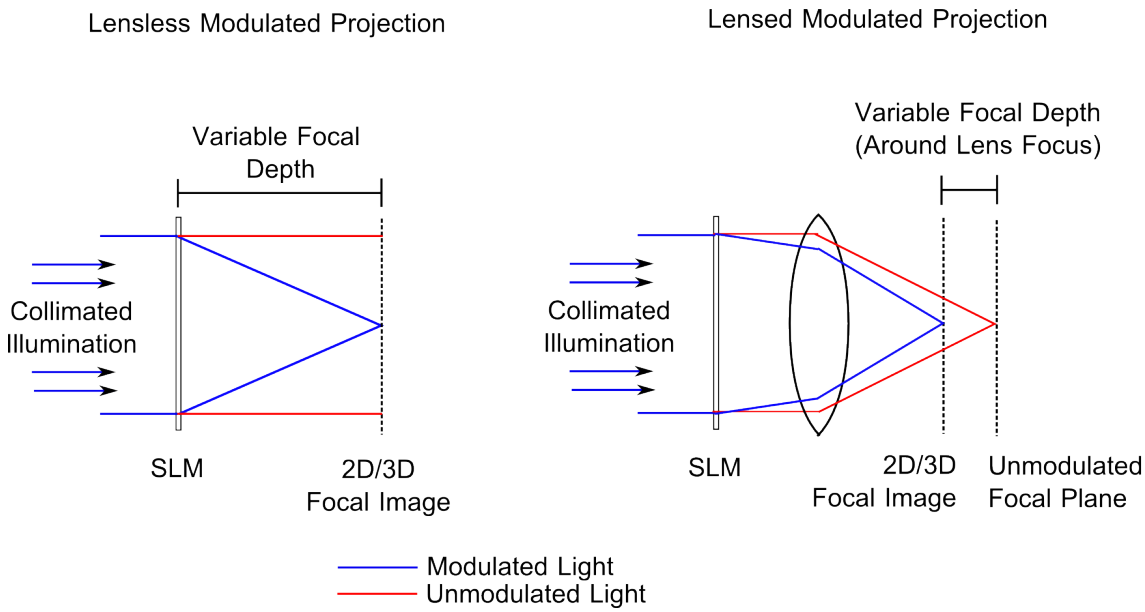


Figure 4.3: Diagrams of lens-less and lensed projection showing modulated and unmodulated foci

As discussed in section 5.4.2 when looking to a lensed system we observe large collections of noise at the focal point of the lens resulting in patterns insufficient for exposure. Therefore, in most of the cases in this document we apply a lens-less SLM projection system, where unmodulated energy is distributed evenly leading to a slight loss of contrast rather than an overexposed region. Lensless projection is also of particular interest to us as it is an optical system close to traditional proximity exposure systems, though potentially with a much larger projection gap.

4.6 Substrate alignment

For 3D projection systems, alignment between substrate and mask becomes a full six degrees of freedom problem (x, y, z position and yaw, pitch, roll rotation), whereas in simple mask alignment systems it is often restricted to three or even one degree(s) of freedom (z projection gap only). This significantly complicates the process of alignment. We have developed some basic approaches to solve this issue, however, at this point in time alignment is still a largely iterative problem; where exposures are attempted and the position of the substrate refined until a focal point is found. While we are aware that this issue will need some significant improvement to form a suitable industrial process, we are confident that such modifications will be possible as the technology matures. Our basic alignment process has involved using the active modulation device to project alignment dot patterns onto a substrate. These are viewed through a long working distance microscope system as shown in figure

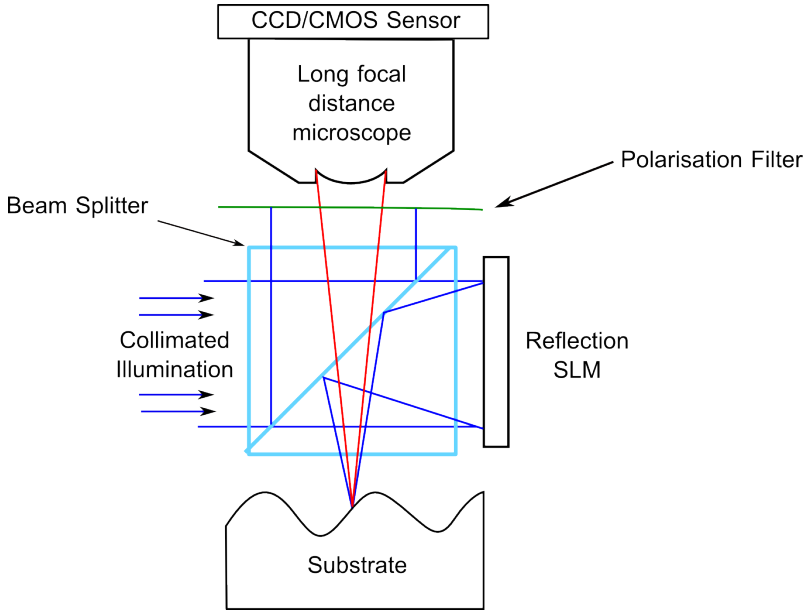


Figure 4.4: Optical alignment setup. First reflection from laser input is blocked from microscope objective by polarisation filter. Reflected/scattered energy from substrate passes through filter due to change in polarisation allowing focal spots to become visible.

4.4.

This process allows us to find the absolute distances between the hologram and substrate at a number of points without full exposure of the substrate. From this we can iteratively adjust the position and angle of the substrate until alignment is accurate for all focal points.

4.7 Conclusion

This chapter has explored the known concepts behind holographic optical projection and how they apply to lithography and holography. A basic set of system constraints has been settled upon and its implications examined. We constrict ourselves largely to a laser-based, spatially-filtered optical source - namely a Coherent CUBE 405nm laser. This allows us to make full use of the coherence of the system to project high resolution 3D patterns. This comes at the expense of some issues with noise which will be addressed in 5.12. An active SLM modulator is to be used to project our phase only holograms, specifically a 'Holoeye' PLUTO 1080p liquid crystal modulator. The phase only constraint maximises diffraction efficiency and provides a simple constraint which is easily compatible with many of the iterative algorithms described in chapter 3. Post-modulation optics have been dismissed for most of our applications as lenses restrict the range of depth available to the modulator, which

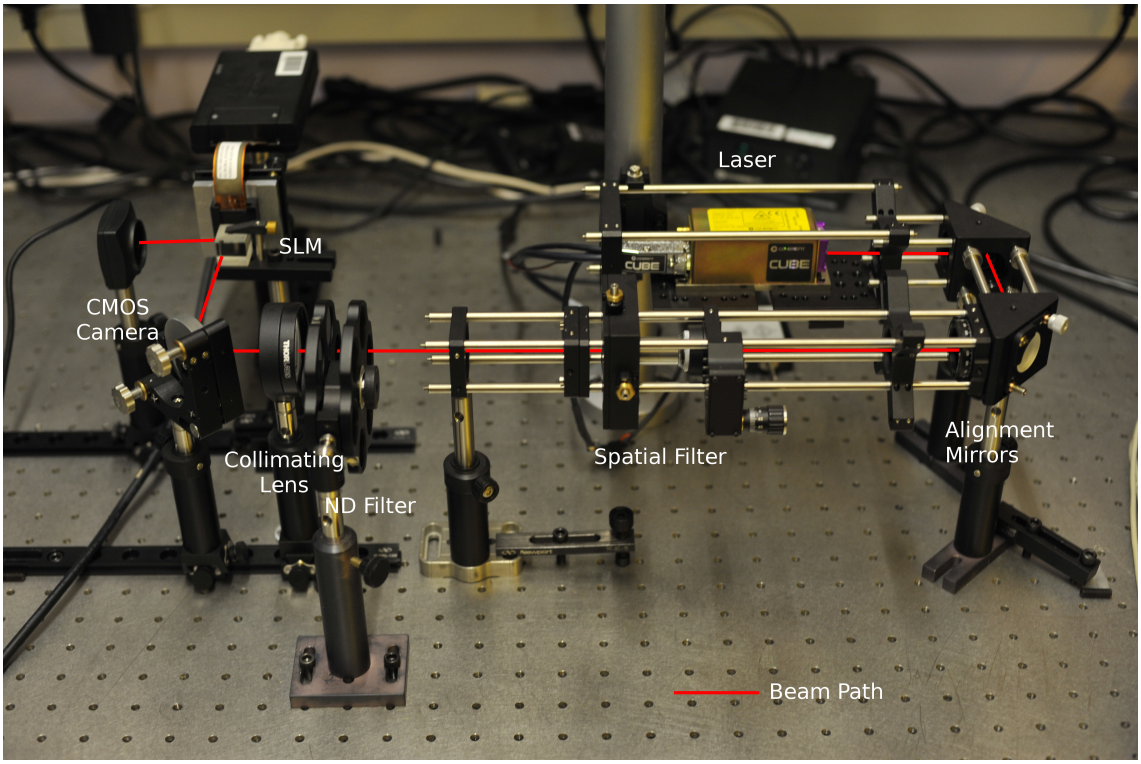


Figure 4.5: Photograph of the practical optical system set-up.

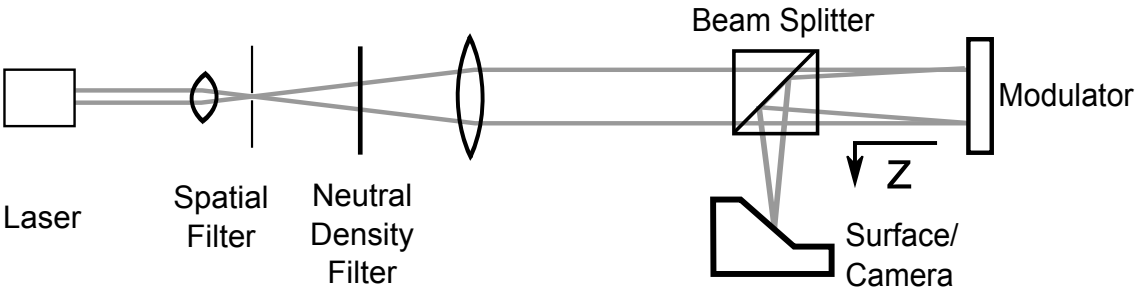


Figure 4.6: Schematic of optical set-up.

needs to be able to focus at a distance large enough to account for a beam splitter. An alignment system has been discussed which has been implemented to some effect to create the practical lithographic demonstrators shown in the final sections of this document.

An image of the practical optical system used for our experimental results can be seen in figure 4.5, with its schematic seen in figure 4.6.

Chapter 5

Planar Iterative Hologram Optimisation, Limitations and Modifications

5.1 Introduction

We have previously discussed the various implementations of iterative algorithms and their advantages and disadvantages in Chapter 3. We have also discussed a practical setup and its implementation. In this chapter we will outline a planar iterative methodology and investigate its parameters and modifications in simulation, and experiment.

This chapter initially defines and covers sampling issues in optical simulation. We then move on to show examples of implementations of both the basic GS algorithm and a derivative version utilising the AS propagation kernel, which we have adopted and adapted to create patterns for lithography. From these examples we derive typical features of these iterative algorithms for the design of holograms. Moving on, we assess how we might use modifications extending the range of depth propagation to extend the usefulness of the iterative approach using the AS transfer function and examine the consequences of these modifications in terms of our optimisation and quality criteria. We then apply the sampling criteria derived for optical simulation to the iterative scheme and examine its effect on a test pattern. Examining other parameters of the hologram and image we see how they affect our quality metric. These simulation parameters include hologram size compared to image size and bandwidth, image feature size compared to PSF width and background amplitude value. Finally, we go on to discuss strategies for removing coherent 'speckle-like' noise from our iterated phase-only holograms by signal averaging and coherence control, and

show an example of an iterated planar bus pattern in simulation, compared to its experimental recording.

5.2 Resolution and Bandwidth in Optical Simulation

A linear optical system is defined by the shape of its point spread function. This function is determined by the transfer function of the system and is usually no more complex than a filter of fixed size and shape determined by a limiting aperture at a distance of a focal image. In a sampling context, this has important consequences to the simulation of optical transforms. It determines the minimum acceptable resolution to form a complete representation of the image. For reasons of simplicity, it can be tempting to choose the pixel pitch of the modulating device as the simulation pitch when applying a hologram to a fixed pixel-pitch device. However, this can easily lead to sub-Nyquist-rate sampling in an image and, therefore, inaccurate results.

5.2.1 Sampling of an image field in 2D

When determining the simulation criteria for a given application, separate conditions must be taken into account. As discussed in Chapter 2 it is important that the transfer function of a given propagation is taken into account. We have not yet discussed sampling of the image itself. From basic optical principles, it is possible to estimate the size of the point spread function of a given optical focusing element. In many cases this would be a lens with a given aperture, but in this research we look to the size of the hologram.

The feature size for a given aperture size can be estimated from the bandwidth of the optical system, which is defined by its numerical aperture. The Rayleigh criterion is often used as an estimate which derives a resolution criterion from the distance of the peak to the first peak of the 'Airy' discs surrounding the peak. We chose to use a more signal-processing-friendly measure in which we ignore the constants imposed by many resolution approximations and merely state that the resolution is the inverse of the optical bandwidth i.e.

$$\theta = \tan^{-1}\left(\frac{r}{z}\right) \qquad NA = n\sin\theta$$
$$\nu_{\max} = \frac{2NA}{\lambda} \qquad \delta x = \frac{1}{\nu_{\max}} \tag{5.1}$$

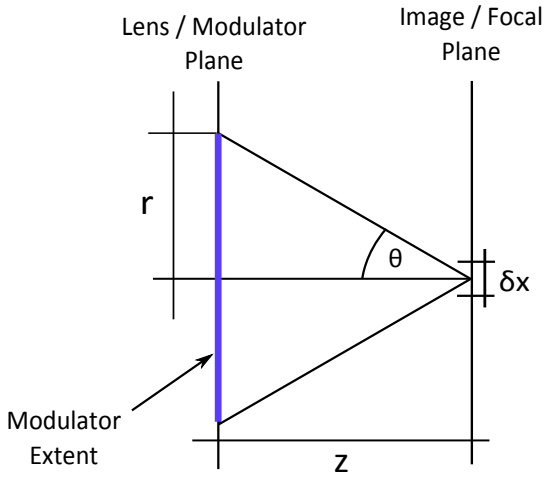


Figure 5.1: Optical system parameters schematic.

Where θ is aperture angle, NA is numerical aperture, λ is light wavelength, $\nu_{\max}$ is the maximum signal frequency in the optical system, and δx is our expected feature size (as shown in figure 5.1). Note the factor of two when calculating the maximum frequency of the optical system is added to account for the difference between 'amplitude' and 'intensity.' The maximum frequency of an intensity pattern is twice that of the maximum frequency of the underlying amplitude field [17]. This is understood when we know that the multiplication of an amplitude field with its conjugate to obtain the intensity pattern, correlates to an auto-convolution of the spectrum of the field which means doubling its maximum frequency. It is sufficient to sample all amplitude fields at half this rate and up-sample the pattern when moving to an intensity analysis, but for simplicity we will always attempt to obey this more stringent intensity sampling criterion.

In most cases considered in this thesis we have used a square hologram for convenience. This is not ideal when considering this optical theory as a circular aperture is often assumed. However, it is possible to consider the worst case aperture (i.e. smallest aperture, lowest resolution, but lower sampling rate required) as well as the best case scenario (largest aperture, highest resolution, and higher sample rate required) for a circular aperture and calculate a sampling criterion based on these. Graphically this is shown in Fig5.2.

As an example, a for a square hologram with width 4.096mm, Fig5.3 shows the

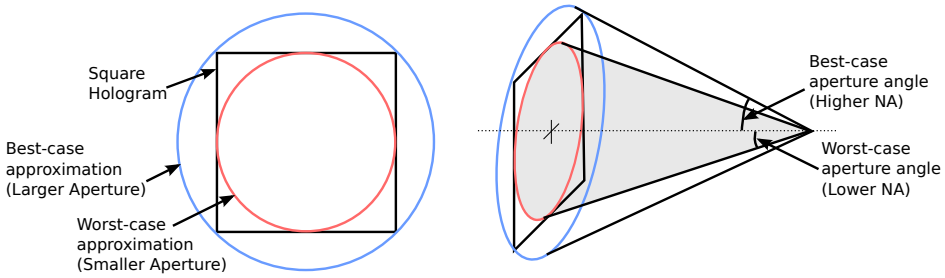


Figure 5.2: Hologram aperture approximation showing assumptions for best-case and worst case numerical aperture

upper and lower analytical limits for a range of focal distances. Shown between the limits is the width (full width at half maximum, FWHM) of a simulated zone plate focal point intensity, which demonstrates the ultimate practical resolution of an optical system. For an optical simulation the field is ideally sampled at the pitch of the PSF (i.e. the Nyquist rate) at most.

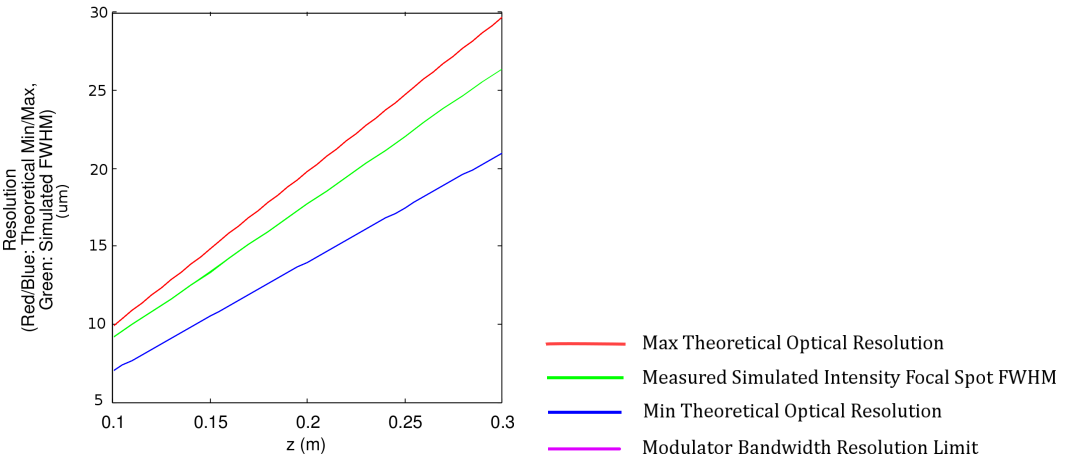


Figure 5.3: Theoretical resolution limits compared with simulated zone plate point image full width at half maximum (FWHM).

5.2.2 Error Measures

When considering the GS approach the initial error measure to be considered is the Sum of Squared Error. This is justified by the idea that this measure can only decrease from one iteration to the next. To calculate the SSE we take the amplitude of the ideal pattern and the amplitude of the simulated image field, take the sum of the square of the difference and divide by the sum of the square of the ideal pattern. This gives us an absolute measure of the error in their intensity patterns.

$$E = \frac{\sum(|A_{ideal}| - |A_{actual}|)^2}{\sum |A_{ideal}|.^2} \tag{5.2}$$

This however, is a rather poor measure of error for lithographic patterns. As it becomes apparent when using the iterative approach with a coherent laser source, reducing the SSE does not necessarily imply the removal of defects from the lithographically exposed pattern. Therefore to shed more light on the suitability of a hologram and image for lithography we turn to more rigorous analysis metrics which allow us to discuss the uniformity and potential flexibility of a pattern in an exposure context.

A useful and well known measure of image contrast consists of the following:

$$C = \frac{H - L}{H + L} \tag{5.3}$$

Where H designates a high value, and L designates a low value.

This kind of contrast metric is used by Wong [6] in the context of lithography as a baseline for estimating the viability of a lithographic pattern, which is why we adopt it here.

These High and Low values are usually set to the extremes of the intensity pattern being considered, but given the variation apparent in many of the holograms generated, this too would be an unreliable estimate of the patterns viability. We chose to use this contrast measure in a more strict way, in order to give us an idea of the gap between the low region and the high region.

Binary images used as amplitude constraints are also divided into high and low regions. Taking this as a mask, and applying it to the image field for assessment, we can choose the highest value in the low region and the lowest value in the high regions of the image. The larger the contrast value calculated by this method with this modification, the larger the gap between the high and low region with respect to the offset of the gap (i.e. the average value), and the easier it becomes to expose an image and get the timing correct so that the exposure threshold is between these two limits. This gives a continuous and correctly-shaped pattern.

This contrast measure has other features which can be considered useful for lithography. It is scale invariant, meaning that varying a source intensity through a fixed hologram is likely not to have an effect on the contrast measure as long as the image intensity pattern scales with input energy. In addition this measure penalises positive offset i.e. if a constant value is added to both H and L (as is potentially the case with unmodulated light shining through the hologram) we observe a decrease in this contrast parameter. This contrast measure is particularly appropriate when

assessing the pattern of continuity in our binary image regions.

5.2.3 Bandwidth Limits in Simulation and Experiment

The bandwidth of an optical system can be considered as the frequency domain equivalent of the resolution of a system. It is possible and sometimes quite useful to think about the limit on the size of a hologram (or the size of any aperture) as a limit on the bandwidth of the signal at its focal plane. This approach can lead to the derivation of the usual equations of resolution based on numerical aperture which we have used to estimate system resolution above. We will refer to this bandwidth limit as the *optical* bandwidth limit.

The SLM of any modulation device has a fixed and finite pixel size, leading to a limit in signal bandwidth. We will refer to this as the *modulator* bandwidth. This limit implies that if we try to implement a hologram which exceeds this physical sampled bandwidth, the re-sampled implementation of this hologram may contain an aliased signal.

If *modulator* bandwidth is greater than *optical* bandwidth the optical system, i.e. aperture, is the limit on resolution. This can be made clear by considering the Ewald sphere and a representation of the optical system in frequency space as in figure 5.4. The Ewald sphere (or circle projected into 2D, as shown here) can be created from the parameters of an optical system.

The wavelength of the light used (λ) with the refractive index of the propagation medium (n) is used to draw a sphere about zero in frequency space (radius = n/λ). Each point on the sphere represents a plane wave travelling in a given direction (and therefore the whole sphere, all possible plane waves in a 3D field with wavelength λ). Secondly, this sphere is then limited to a conic section by the aperture angle, which limits us to a particular bandwidth in ν_x,ν_y and ν_z. The diagram in figure 5.4 shows the optical bandwidth of a system determined by the aperture angle at the focal image depth. Thirdly, the bandwidth in the modulator is limited by the sample pitch of the device in the hologram plane. The limiting factor may be the optical bandwidth or the modulator bandwidth, depending on focal depth (z).

We can examine what happens when the optical bandwidth exceeds the modulator bandwidth by considering a zone plate implemented on a phase-only modulator. We implemented a simulation of a fixed 8μm modulator by generating a simulation with a much higher resolution (2μm in this case). We then simulated the larger pixel size of the modulator by a re-sampling function wherein each 4 by 4 group of pixels was assigned the value of the pixel in its upper left, as shown in figure 5.5. This allowed us to simulate the pixel pitch of the modulator without forgoing

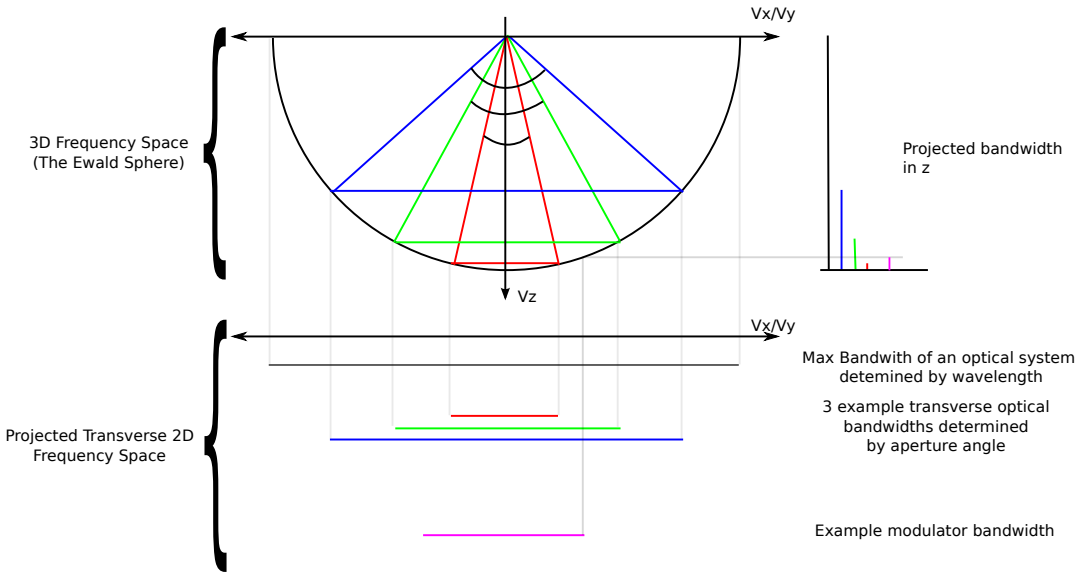


Figure 5.4: Red/Blue/Green: Examples of transverse/longitudinal optical bandwidth extent in the frequency domain (BW) determined by construction of an Ewald sphere with radius λ/n, Pink: Potential modulator transverse BW showing a projected longitudinal BW. The modulator BW would be the limiting factor in resolution for the case of an optical system with the Green or Blue optical BWs, but in the case of the system described by the Red bandwidth the opposite is true and the optical bandwidth would limit the resolution of this system.

resolution in the sampled image plane, by using an AS propagation kernel and a convolution approach. With this method, input and output plane sample spacings remain constant.

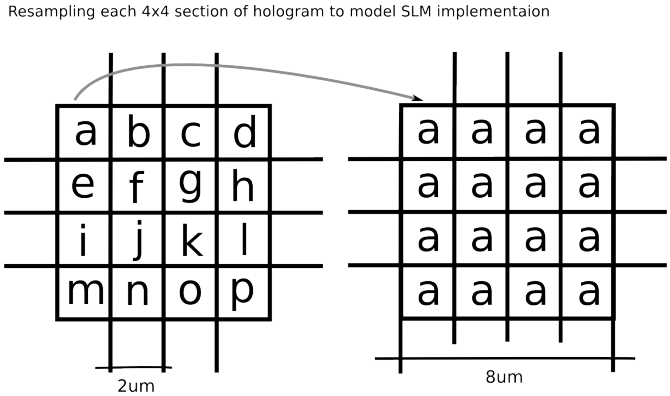


Figure 5.5: Re sampling function for lower resolution modulator from higher resolution simulation

Figure 5.6 shows the FWHM image point size measurement for zone plates with a varying z (and therefore varying optical bandwidth). It can be seen that where the analytical resolution would exceed the limit imposed by the bandwidth of the modulator (approx $4\mu m$) the practical implementation re-sampling causes erratic

deviation from this ideal.

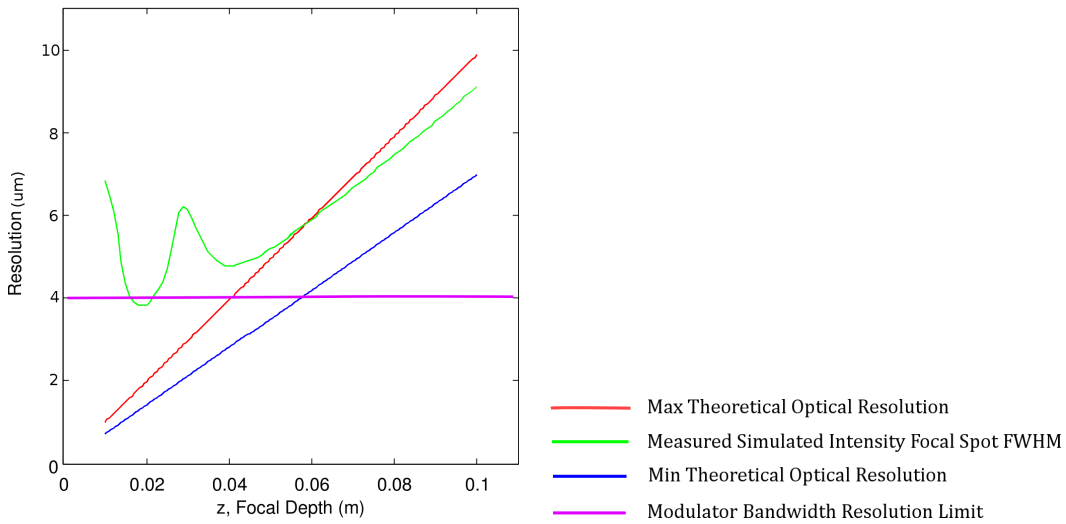


Figure 5.6: Modulator bandwidth limit effects when applying re-sampling

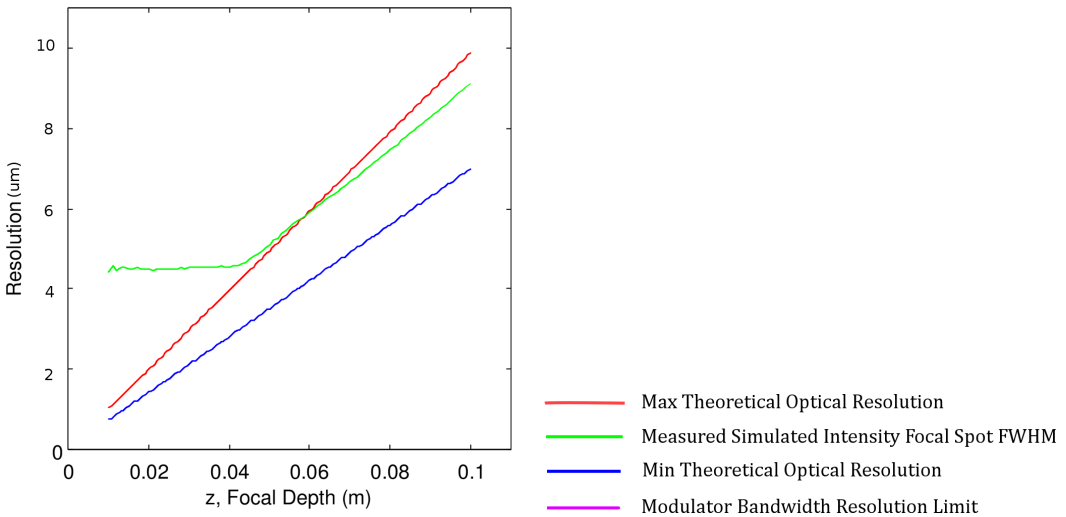


Figure 5.7: Modulator bandwidth limit effects when applying filter and re-sample at modulator bandwidth

To address this, it is clear that the naive re-sampling functions are not enough to ensure that the hologram is correctly implemented when the image is in close proximity to the modulator (i.e. when the size and sampling constraints exceed that of the modulator). To compensate, the hologram can be filtered such that the analytical pattern bandwidth cannot exceed that of the modulator, before re-sampling the hologram for implementation. Fig5.7 shows how the width of the PSF behaves when brought close to the modulator under these filtered conditions, i.e. the variation below this limit reduces significantly. It should be noted that variation still

exists, which is likely because the hologram is re-quantised to phase-only after being filtered. An iterative approach could address this, but it is relatively unimportant in this case.

5.2.4 Resolution and Bandwidth Conclusions

In this section we have defined useful sampling limits with regards to optical simulation and experiment. The consequences of these limits with regards to iterative algorithms are further explored in section 5.8.1. Furthermore, we have examined the effect of pixelation and filtering of the field with regards to a zone plate. This pixelation models the implementation on a pixelated SLM and has important consequences. The pixelation routine is found to be problematic when considering lower bandwidth modulators. A filter-and-pixelate approach is defined, which removes some of the variance associated with applying a high resolution pattern to a real modulator.

5.3 Planar Iteration Scheme

5.3.1 Gerchberg Saxton Method Outline

The behaviour of the GS iterative scheme was laid out in the original paper [35]. An overview is described in section 3.3.2. A basic code implementation of the FFT algorithm described can be seen in listing 5.1

Listing 5.1: An implementation of a simple FFT based iterative routine

```
%% Generic 2D iteration function
function ...
    [output,focal_image,E]=quick_iterate_FFT(n_iter,image,SEED_HOLO)
    % Data Initialisations
    HOLO=SEED_HOLO;
    P=HOLO;

    for n=1:n_iter % Iteration Loop

        P(:,:)=fft2(P(:,:)); % Forward Transform

        focal_image = abs(P); % Cache Image

        P=(P./abs(P)).*image; % Image Constraint
        P(:,:)=ifft2(P(:,:)); % Reverse Transform
        P(:,:)=((P./abs(P))); % Hologram Constraint

        HOLO=P(:,:); % Cache Hologram

        E(n)=SSE(image./max2d(image),abs(P)./max2d(abs(P))); % ...
            Error Calculation
    end

    % Set Outputs
    output=HOLO;
    focal_image = focal_image.^2;
end
```

Using the above function we have set up a test which loads in a picture of the Durham University crest into an upper quadrant, which is then optimised using this GS algorithm implementation. Figure 5.8 shows the image created in simulation, the hologram created and the image pattern from an experimental implementation.

This test example gives initially promising results in simulation which do not

Simulation Width/Height	SLM Pixel Size	Number of Iterations
1024 pixels	$8\mu m$	50

Table 5.1: Parameters for simulation and experimental reconstruction shown in figure 5.8

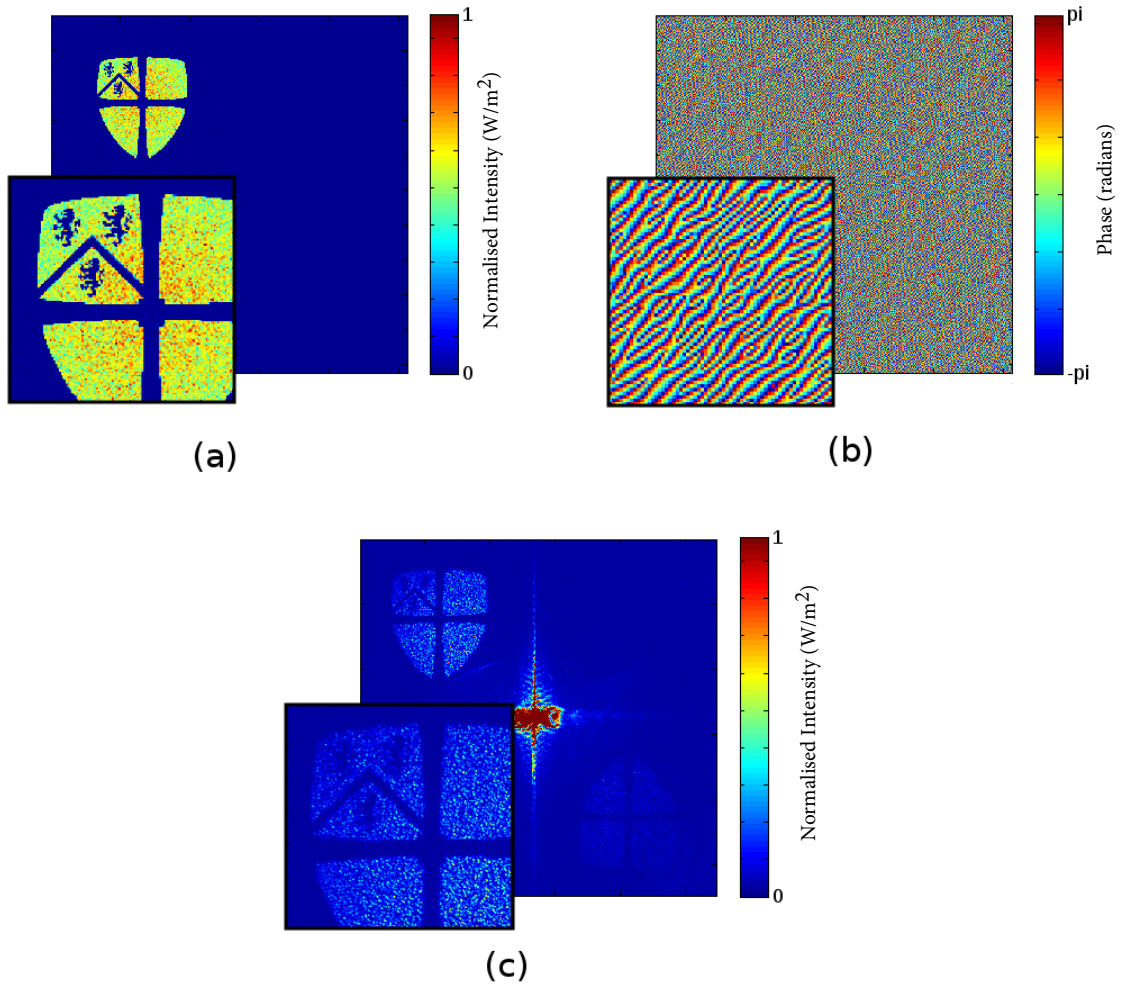


Figure 5.8: Simple Iterative Method Implementation Evaluated Experimentally. (a) Simulated image, (b) Iterated Hologram, (c) Experimentally recorded image.

hold up when tested experimentally. This can be seen in figure 5.8 wherein an iterated simulated image plane showing the Durham crest (a) can be seen with its corresponding Fourier hologram (b). The simulated image clearly does not match that shown in experiment (c); noise levels are much higher and severe vignetting obscures parts of the image. By zero padding the hologram or interpolating the field the noise present in the system becomes visible, however, the accuracy compared to the experiment is still questionable.

Some other troubling features can also be observed. Firstly, the large focal plane spot caused by unmodulated light from the modulation device. If left unfiltered this could easily destroy a lithographic exposure. Secondly, the weaker conjugate image in the lower right quadrant which may produce unwanted exposed patterns. Finally, the non-uniform intensity trailing off towards the edge of the pattern can make getting a uniform exposure much more difficult.

As suggested by the title of "error reduction algorithm" the GS algorithm itself cannot increase the SSE of the pattern and, at worst, stagnates. Many derived methods are reasonably far removed from this original approach, either by use of special transfer functions applied in the Fourier domain or by the approximations used when simulating a propagation.

5.4 System considerations

5.4.1 Image depth

The system of projection we chose (i.e. that of a free space propagation between hologram and image, rather than using a lensed system) is justified by the limitations a lens imposes. With a lens focal point the image becomes centred around this spot and the limited bandwidth of the modulator can move the focal image only from the lens foci in both planar translation and depth. When not using a lens, the minimum focal distance is still limited by the bandwidth of the modulator, but the system remains flexible beyond this minimum focal distance. The trade off occurring here for this wider range of depth is a reduction in minimum feature size. In addition, a lens system can be used to photo-reduce a pattern. This allow far smaller features (albeit with a greater risk of aberration).

5.4.2 Collected Background Noise

In a lensless imaging system unmodulated light reflected from the modulator or other surfaces in the optical system (such as the beam splitter) remains distributed as general background energy across the whole image field. With a lensed system all unmodulated light is collected into a central point in the image focal region. Depending upon implementation, this could be a serious problem in a lithography system. It would cause an unusable area of the substrate to be exposed. This would require that the image is confined to a quadrant far from the centre of the focal region. By not using a lens, therefore, we gain a potentially larger image region, though we do have to deal with a greater amount of general background noise.

Alternatively, after projection a zero order filter could be implemented in the beam blocking unmodulated light at the focal point of the lens and then reprojecting the filtered image. However, this adds another stage and significant complexity to the optical system.

5.5 Lensless Planar Iteration

Because of the issues above, and as several before us have done [37, 38, 40], we chose to replace the simple Fourier transform propagation with a convolution based transform. This simulates a lensless system by adopting an appropriate transfer function between constraint planes. Largely through this work we have chosen to use the AS transfer function method to propagate between parallel planes in the image and hologram.

This decision is justified by the nature of the numerical evaluation. We are able to use the same sample rate on both sides of the propagation (unlike the basic numerical Fresnel method), therefore the transform remains relatively computationally efficient (as compared to the RS method). Finally, the method uses minimal approximations (in contrast to the Fresnel kernel convolution). For a more in depth comparison of these methods please refer to chapter 2.

The method can then be described as follows, and is outlined in figure 5.9.

1. Initialise a random phase field. $U(x, y) = \text{Rand}(x, y)$. Where $U(x, y)$ represents our sampled field in the hologram plane.

2. Multiply field angles by a hologram amplitude constraint pattern (Top hat function for phase-only). $U'(x, y) = \text{Angle}(U(x, y)) \times C_H(x, y)$, where C_H is a sampled top hat amplitude function

3. Propagate forward to an image plane by convolution with transfer function (h) for a given z distance. $I_1(x, y) = U'(x, y) * h(x, y, z_1)$

4. Apply an image amplitude constraint pattern. $I'(x, y) = \text{Angle}(I(x, y)) \times C_{I1}(x, y)$ Where C_{I1} is the constraint pattern to be applied on the 1^{st} constraint plane.

5. Propagate back to the hologram plane. $U(x, y) = I'(x, y) * h(x, y, -z_3)$

6. Repeat process from step (2).

Figure 5.10 outlines the basis of this method schematically.

The code listing 5.2 shows an implementation of this approach in a Matlab function. The parameters passed to this function are mostly self-explanatory (z, lambda, x_samples, y_samples, holo_size_x, holo_size_y, sample_pitch). The 'n_iter' variable determines the number of iterations to be executed, the 'image' variable is a matrix containing a sampled amplitude image in 2D with high values set to one and background values set to some small value usually in the range 0 to 0.2.

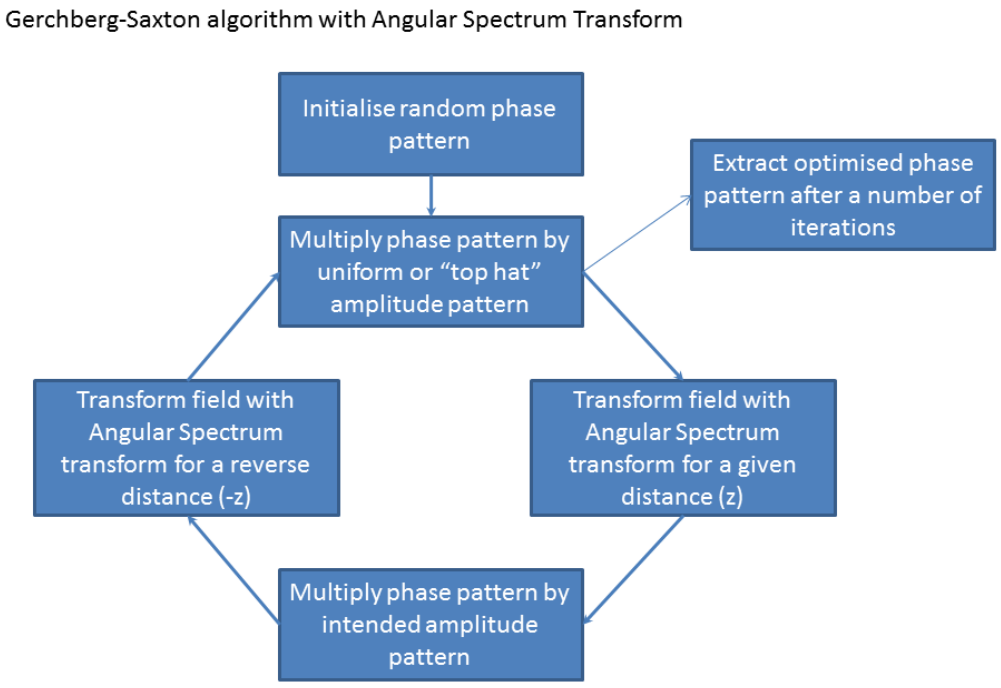


Figure 5.9: Flow diagram outlining the key steps in an iterative algorithm using the angular spectrum transform

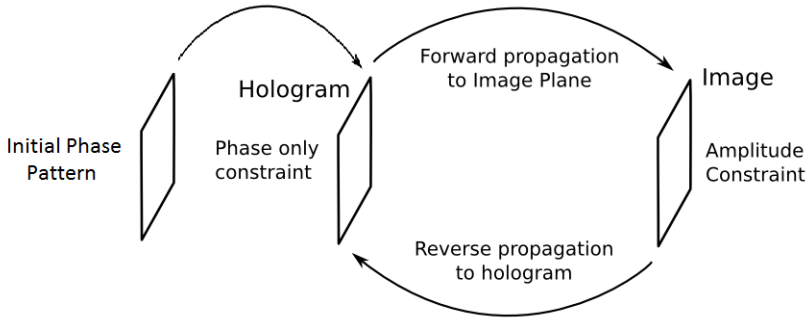


Figure 5.10: A planar iterative scheme illustrated graphically

The 'SEED_HOLO' argument here is an initial condition for the hologram phase pattern, again in a sampled matrix of complex values. Finally the 'prop' value passed in determines the size of a circular filter to be applied to the simulation transfer function who's radius is a proportion of the size of the sampled field. The use of this bandlimit is discussed in section 5.8.1, but for our initial simulations it is set to $\sqrt{2}$ which indicates that the simulation will not be filtered in this way, instead relying on the simulated optical bandwidth limit caused by the size of the hologram

Listing 5.2: An implementation of a Angular Spectrum based iterative routine

```
26 %% Generic 2D iteration function
27 function [output,focal_image,E,C] ...
       =planar_iterate(z,lambda,x_samples,y_samples, ...
       holo_size_x,holo_size_y,sample_pitch,n_iter, ...
       image,SEED_HOLO,prop)
28
29      %% Pre-Calculate Filtered Transfer Functions
30      mask=fftshift(gen_bl_mask_elip(x_samples,y_samples,
31              1/(x_samples/2),1/(y_samples/2),prop));
32
33      H_back    = calc_AS_TF(-z,lambda, ...
           x_samples,y_samples,sample_pitch) .* mask;
34      H_forward = calc_AS_TF(z,lambda, ...
           x_samples,y_samples,sample_pitch) .*  mask;
35
36      %% Data Initialisations
37      E=0;C=0;
38      HOLO=SEED_HOLO;
39      P=HOLO;
40
41      for a=1:n_steps
42          P(:,:)=convolve_with_TF_GPU(bandlimit(P(:,:), ...
               holo_size_x,holo_size_y,0),H_forward(:,:));
43      end
44
45      %% Iterate Loop
46      for n=1:n_iter
47          %% Image Amplitude Constraint
48          P=P./abs(P).*image;
49          %% Propagate To Hologram
50          P(:,:)=convolve_with_TF_GPU(bandlimit(P(:,:), ...
               holo_size_x,holo_size_y,0),H_back(:,:));
51          P=P./abs(P);
52          HOLO=P;
53          %% Propagate to Image
54          for a=1:n_steps
55              P(:,:)=convolve_with_TF_GPU(bandlimit(P(:,:), ...
                   holo_size_x,holo_size_y,0),H_forward(:,:));
56          end
57          E(n)=min_SSE(image,abs(P),0.00001);
58      end
59      % Outputs
60      output=single(HOLO);
61      focal_image=single(abs(P).^2);
62  end
```

to limit the resolution of the simulation.

5.6 Simulation Test Patterns

We test the method described above by attempting to optimise a hologram for a variety of different shape patterns (a line, a bus of lines, a cross and a square), for one fixed set of optical and simulation parameters. From this we are able to discern what kinds of features we would expect to see in an optimised focal image.

5.6.1 Setup and Parameters

It is important to note that the simulation parameters chosen here are selected to avoid the issues of optical bandwidth (as discussed in section 5.2 and further addressed in section 5.8.1). In this case, the hologram is forced to be small enough such that the bandwidth of the simulation is sufficiently larger than the optical bandwidth. Furthermore, the optical bandwidth is limited in such a way that the bandwidth of the modulation device is sufficient. The z distance is chosen so that the propagation transfer function does not alias, whilst also maintaining these other sampling issues as this can cause problems which are understood and addressed in section 5.7.

Note that the simulation size is much larger than the image size to accommodate a suitable representation of the transfer function spectrum without having to resort to filtering. The calculated minimum image feature size from the optical parameters is $21\mu m$. As such, a sample size of $3\mu m$ is more than sufficient. Intended feature size is limited to a value of 2 pixels or $6\mu m$ which should leave us with features no smaller than the PSF of the system. Constraint values in high and low image regions are set to reasonably-scaled non-zero values, though it should be noted that this issue is further discussed in section 5.11.

From these parameters and the equations derived in section 5.7.1 we can calculate the maximum propagation distance without aliasing of the transfer function is

Propagation distance (z)	Wavelength (λ)	Sample/Pixel Pitch
$4cm$	$405nm$	$3\mu m$
Image Width/Height	**Simulation Width/Height**	**Feature Width**
256 pixels	4096 pixels	2 pixels
Constraint Low Value	**Constraint High Value**	**Number of Iterations**
0.1	1.0	100

Table 5.2: Simulation parameters used throughout the experiments in this chapter, except where otherwise specified.

4.5cm. Therefore, our value of 4cm depth propagation is suitable. Later we will see how this limit can be broken by modification with little consequence.

Seed Pattern Generation

The following set of planar iterative tests have been performed three times, each with a different seed pattern. In the classical GS method paper it was noted that the seed pattern was required to be non-uniform and non-zero. For the iterative planar algorithm (using either the AS or Fresnel transfer functions) we find that this is not the case. Therefore, we tested each with a random phase pattern, uniform unit phase pattern, and finally a single simple analytical line hologram approximation.

5.6.2 Single Line Test Pattern

Figure 5.11 shows an analysis of the analytical line pattern used as a seed for the following iterative experiment. In this case the hologram is small compared to the

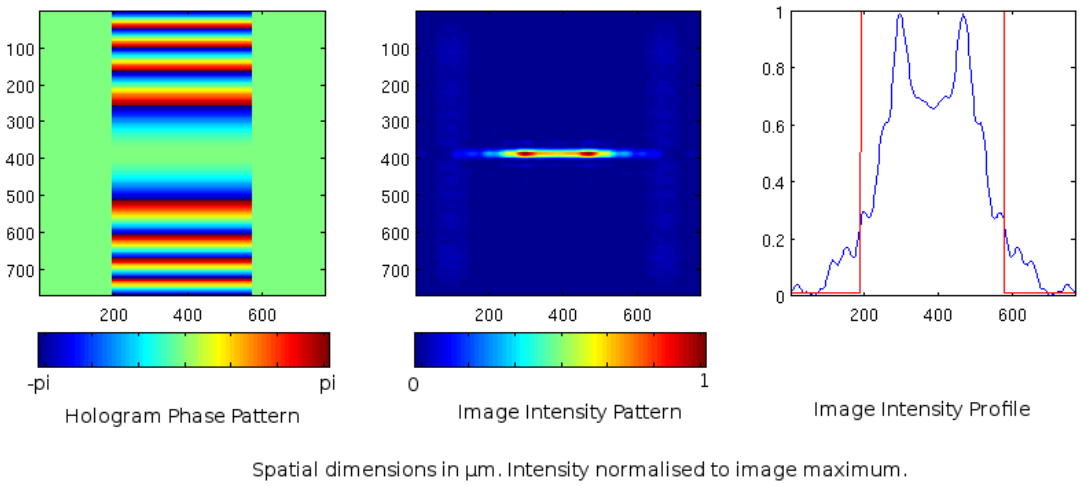


Figure 5.11: Analytically derived single line seed hologram and its simulated image intensity pattern, with central profile.

focal depth of the image, hence only a few fringes are visible in the hologram. The intensity pattern and profile of this analytical approximation show a reasonably good cross sectional profile, but also severe peaks and troughs as well as slow tapering line termination, and a significant change in line length compared to the ideal image pattern (shown in red). Figure 5.12 shows the holograms, image intensity patterns and profiles for a planar iteration experiment carried out with a single line test image.

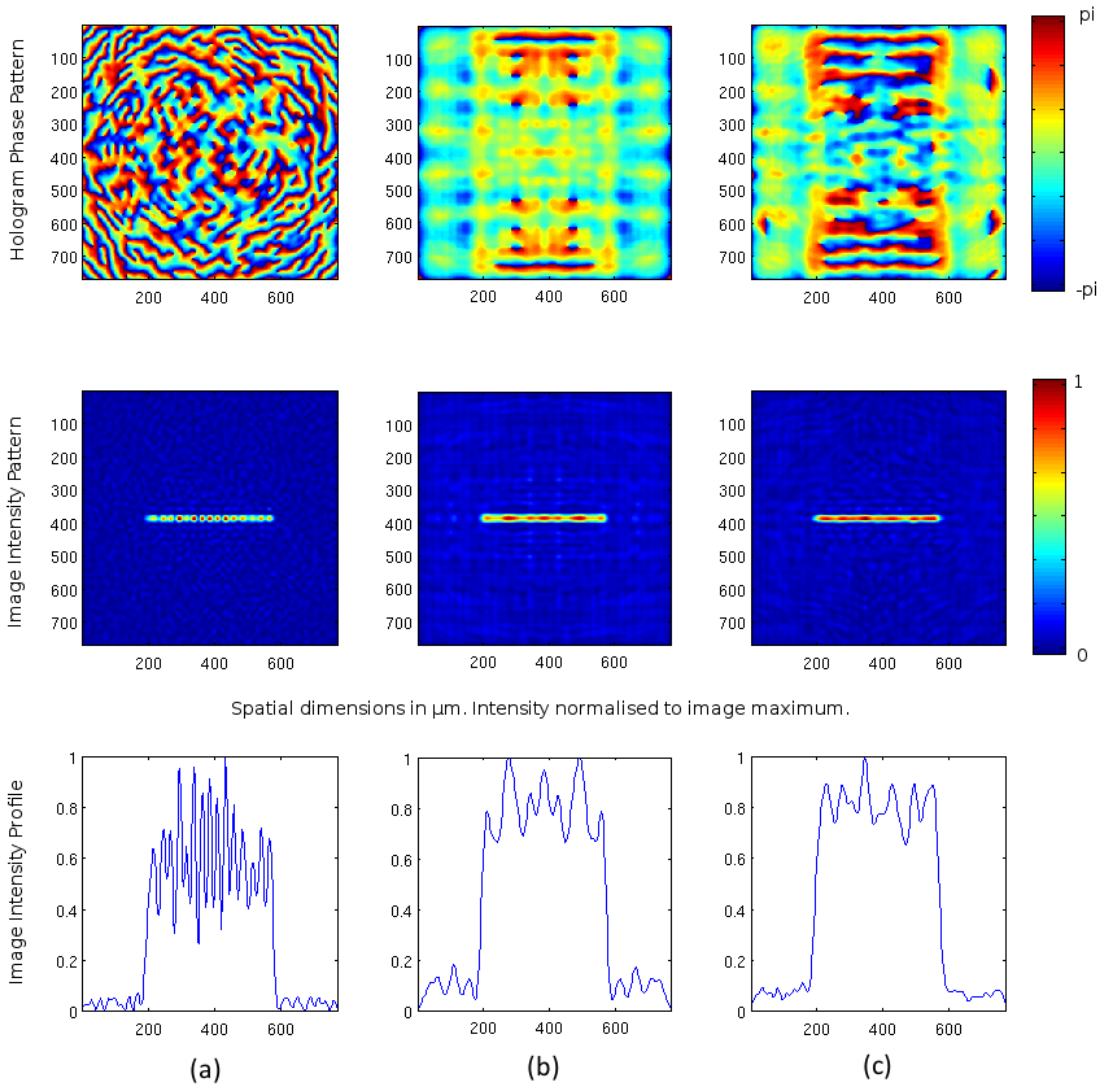


Figure 5.12: Single line test pattern simulation results. (Top) Optimised hologram, (Middle) Image intensity pattern, (Bottom) centre profile of intensity, after 100 iterations. Comparison of (a) random, (b) uniform, and (c) analytical approximation, seeded iteration results.

From this fundamental iterative experiment, certain properties of the iterative algorithm become clear. Although all three cases have arrived at potentially viable solutions ($C > 0$), the random phase seed case still exhibits significant noise in the "high" region, which is not as prevalent in either of the other cases. The hologram shown in (a) indeed also exhibits a more homogeneous speckle-like structure which is not the case for the other variants. All three examples demonstrate a good suppression of noise outside of the "high" image regions, though the structure of the noise does vary. Structural noise "hotspots" are just visible in (b), which impact its

calculated contrast.

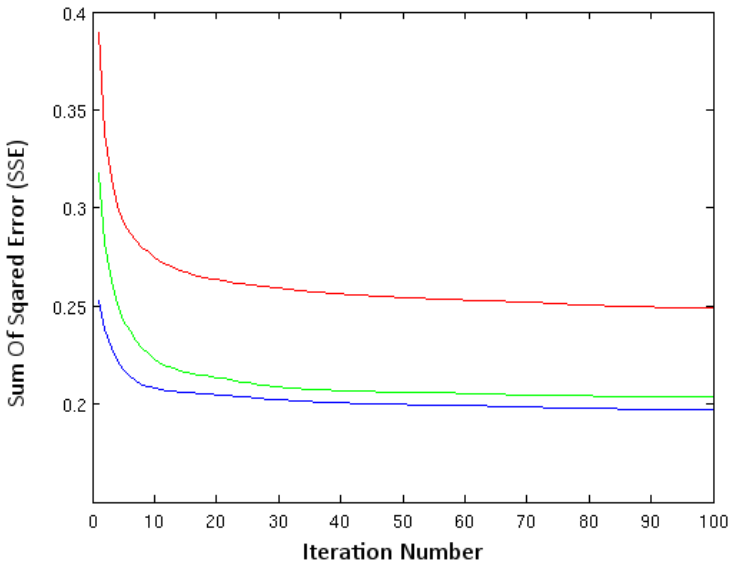


Figure 5.13: Single line test pattern simulation results. Comparison of scale minimised SSE versus iteration number for three different seed phase patterns. (Red) Random, (Blue) Uniform plane phase, (Green) Analytical approximation

Figure 5.13 shows the SSE for each iteration in each seeded cases. This demonstrates that SSE successfully decreases for each of the different seeds, planar and analytical seeds outperform the randomly seeded example by approximately a 0.05 decrease in SSE.

Figure 5.14 shows the measured contrast for the line pattern versus the background noise level. This does not demonstrate the same trend. The analytical approximation outperforms planar and random phases in terms of contrast, but planar phase contrast trends downward after the first few iterations. This results in a final value not significantly different from that of the vastly improved random phase seed.

The analytical approximation starts at a much higher contrast and is yet further improved by the end of the process. It is important, however, to note that the upward trend observed decreases somewhat in the first few iterations before a general upward trend is reached.

Since the analytical pattern starts from such a high contrast, the usefulness of this method could be questioned as showing a negligible improvement. This contrast measure - whilst usefully indicative - is a necessary, but not a complete and sufficient, measure of lithographic quality. In comparison to the simple analytical approximation edge definition is noticeably improved using all three of the seeds,

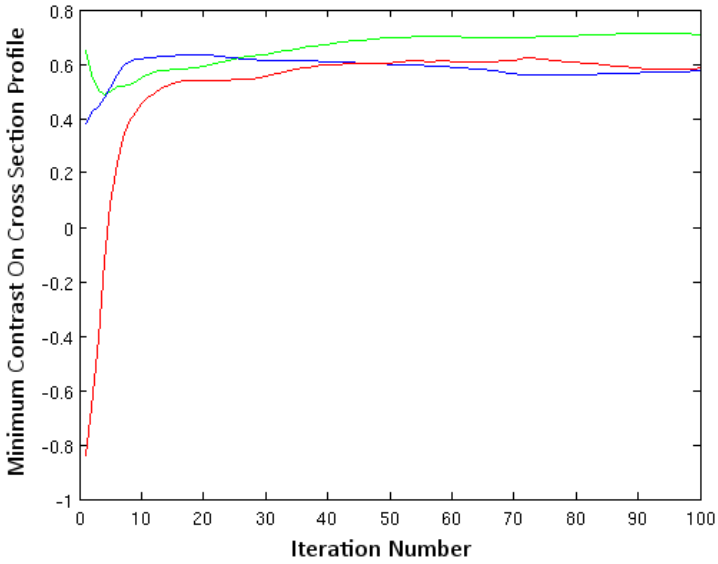


Figure 5.14: Single line test pattern simulation results. Comparison of normalised contrast versus iteration number for three different seed phase patterns. (Red) Random, (Blue) Uniform plane phase, (Green) Analytical approximation

displaying a roll off rate comparable to the width of the PSF of the optical system. Line length in all three cases closely matches that of the intended pattern.

The improved SSE of the system is indicative of an improved diffraction efficiency, i.e. the hologram's ability to push energy into the correct parts of the pattern.

5.6.3 Multi-Line Test Pattern

Figures 5.15, 5.16 and 5.17 show the iterated holograms, intensity patterns and error and contrast evolutions for a multi line bus of eight lines in close proximity.

For this multi-line bus test pattern a single line analytical seed is still applied for comparison with the previous example. Though the central profiles do indeed form suitable continuous regions, it is clear that this is not the case for all of the bus lines. By increasing the complexity of the image, even through this relatively minor change, we have surpassed what can be achieved for this scale of hologram at this bandwidth.

As will be shown in section 5.9 this can potentially be mitigated if we consider extending the size of the hologram up to some limit given a limited bandwidth.

Convergence and stagnation of the SSE to a lower level is similar to previous examples. However, by a combination of contrast measures for each of the line's profiles the overall contrast in the image displays a much more erratic behaviour in all three cases. This is a good example because as previously noted that optimisation

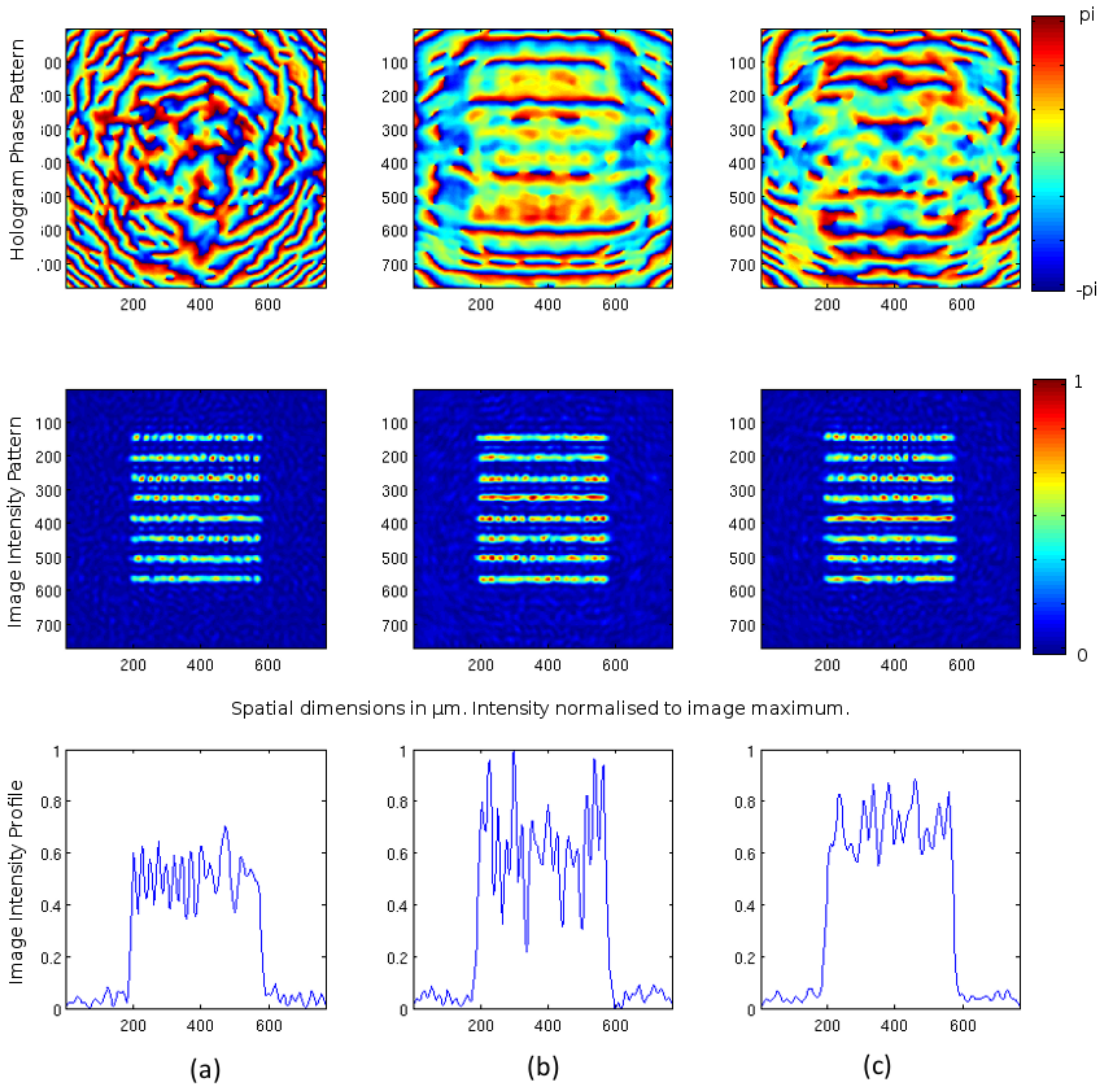


Figure 5.15: Multi-Line test pattern simulation results. (Top) Optimised hologram, (Middle) Image intensity pattern, (Bottom) centre profile of intensity, after 100 iterations. Comparison of (a) random, (b) uniform, and (c) analytical approximation, seeded iteration results.

of SSE does not immediately result in optimisation for lithographic exposure. As shown in the planar and randomly seeded single line examples, the contrast measure is often also improved, however this is not necessarily the case. We do still observe that the iterated pattern results in a very well-defined, if noisy, image in the high regions with edge roll-off comparable to that of the PSF of the optical system.

Figure 5.18 shows a comparison of the line bus pattern for small numbers of iterations (1,5 and 10) to give an indication of the scale of improvement we are observing for this approximately 0.1 reduction in SSE for this particular pattern.

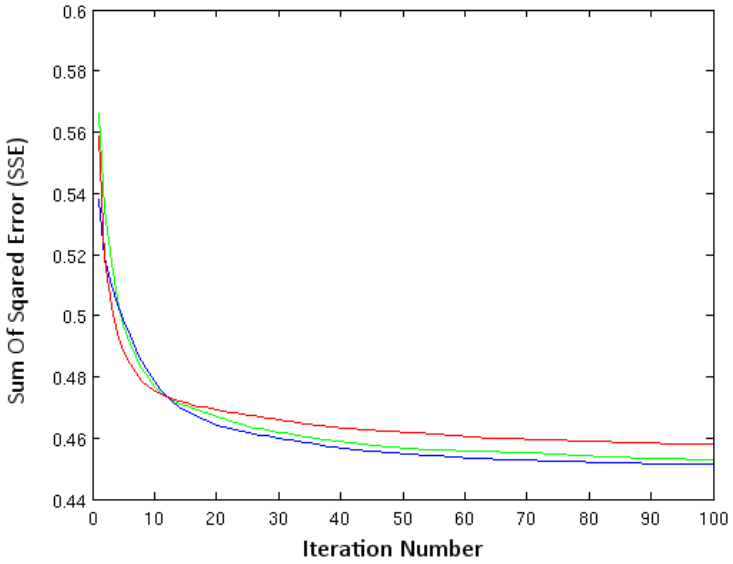


Figure 5.16: Multi-Line test pattern simulation results. Comparison of scale minimised SSE versus iteration number for three different seed phase patterns. (Red) Random, (Blue) Uniform plane phase, (Green) Analytical approximation

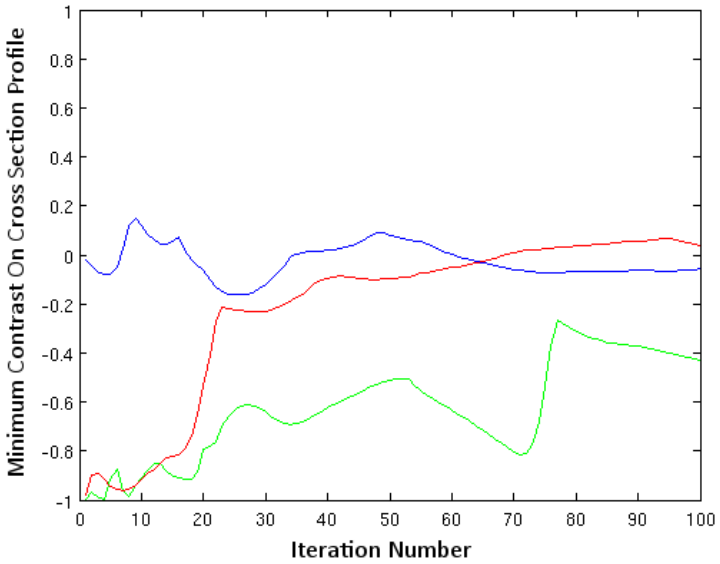


Figure 5.17: Multi-Line test pattern simulation results. Comparison of normalised contrast versus iteration number for three different seed phase patterns. (Red) Random, (Blue) Uniform plane phase, (Green) Analytical approximation

It can be seen in the randomly seeded case that energy is being redistributed in such a way that the pattern fills out over the course of a few iterations. In the plane phase-seeded case the relative background value can be seen to be reduced, improving the average contrast of the pattern. Finally in the seeded case we observe

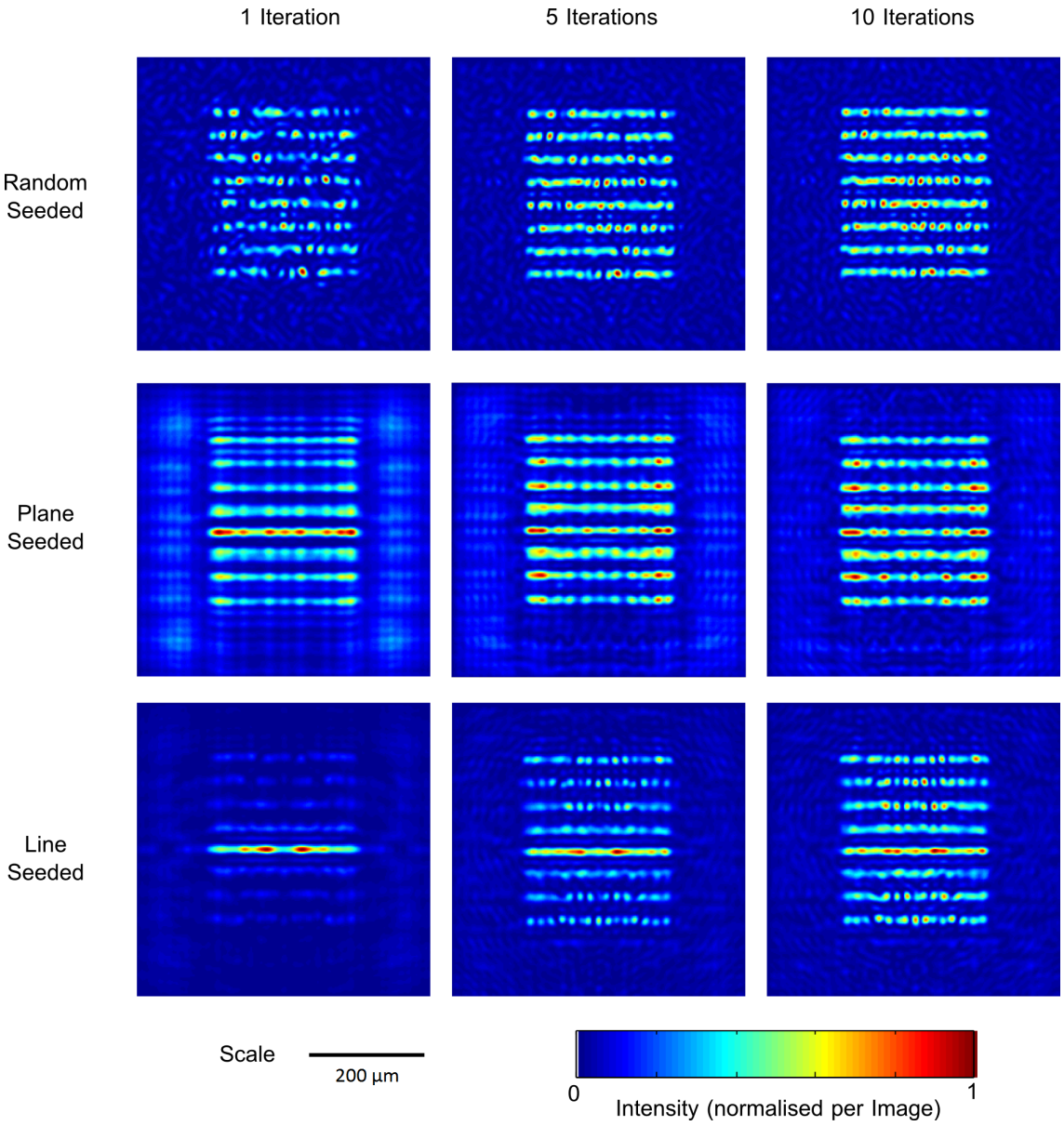


Figure 5.18: A comparison of simulated intensity patterns for small numbers of iterations.

a redistribution of energy from the localised seed line into all of the other lines in the pattern. This gives a good indication that the iterative algorithm is relatively capable of the redistribution of energy in the image pattern. It will be noted later that a more appropriate seed pattern, such as that of a set of lines approximating the image pattern can lead to better results, but this example is a good demonstration of the algorithms flexibility. We do, however, note that in none of the cases are noisy regions optimised away, resulting in poor contrast measurements. Indeed in the plane phase-seeded case, we observe that speckle-like patterns develop from a much smoother initial intensity pattern.

5.6.4 Line Cross Test Pattern

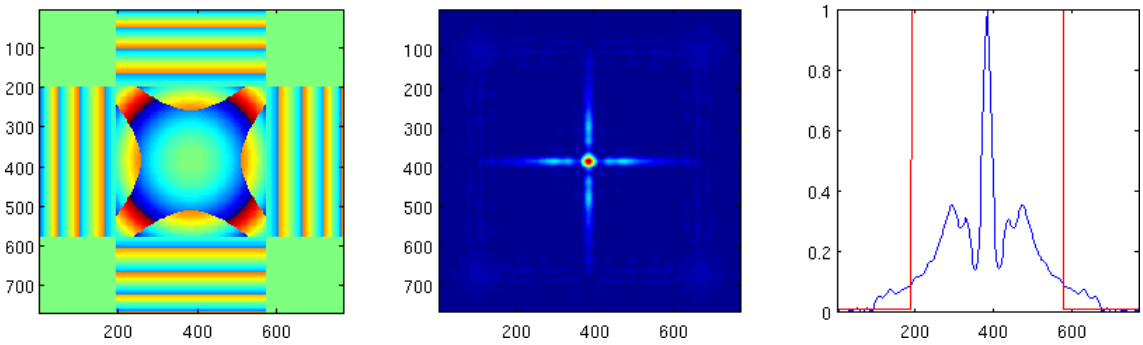


Figure 5.19: Seed pattern used for crossed lines test pattern. (a) hologram, (b) simulated intensity image pattern, (c) central profile of image intensity.

A cross pattern is assessed in figure 5.20, with SSE and contrast evolutions in figures 5.21 and 5.22.

An analytical cross hologram (Figure 5.19) was used in place of the simple line hologram as a seed which demonstrates a further limitation of the analytical seeded approach. The analytical seed image shows that at the centre of the phase constrained analytical cross a large spike in intensity is present with severe nulls surrounding it where the two patterns have interfered constructively and destructively.

The planar iterative algorithm has suppressed some of this spike intensity for the seeded pattern but is unable to completely equalise to a uniform intensity leading to a similar central hotspot. Specific to this example we (McWilliam et al [31], in preparation) can show that for this thin pattern a known phase pattern can be applied which allows a cross pattern such as this to be formed without such a spike.

Therefore this pattern demonstrates a limitation of the algorithm, which is unable to compensate for such a wide variation. It is interesting to note that even in the cases where a seed is not used, starting from a random and a plane phase still results in a peak in intensity at the centre. This suggests that our phase-only constraint restrains the algorithm from removing this central disturbance.

This means that crosses, T-junctions, and certain types of corner can be difficult to pattern using this approach. These are quite serious constraints for lithography because of its commonplace reliance on Manhattan style geometry.

The graphs of varying SSE across iterations show that in this particular case, both seeded and plane phase outperform the randomly seeded case. This is also true for the contrast with the seeded pattern performing best of all.

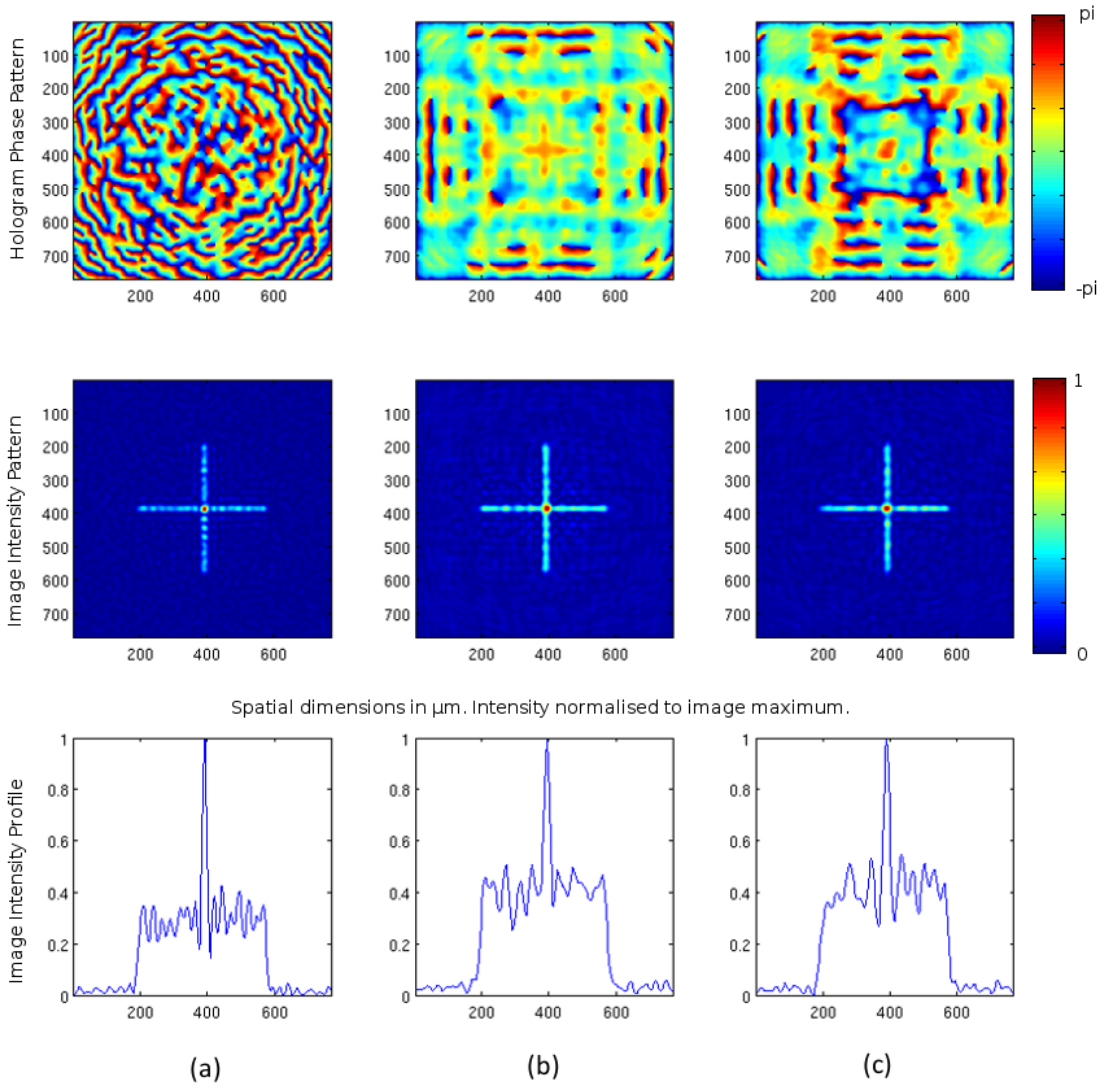


Figure 5.20: Multi-Line test pattern simulation results. (Top) Optimised hologram, (Middle) Image intensity pattern, (Bottom) centre profile of intensity, after 100 iterations. Comparison of (a) random, (b) uniform, and (c) analytical approximation, seeded iteration results.

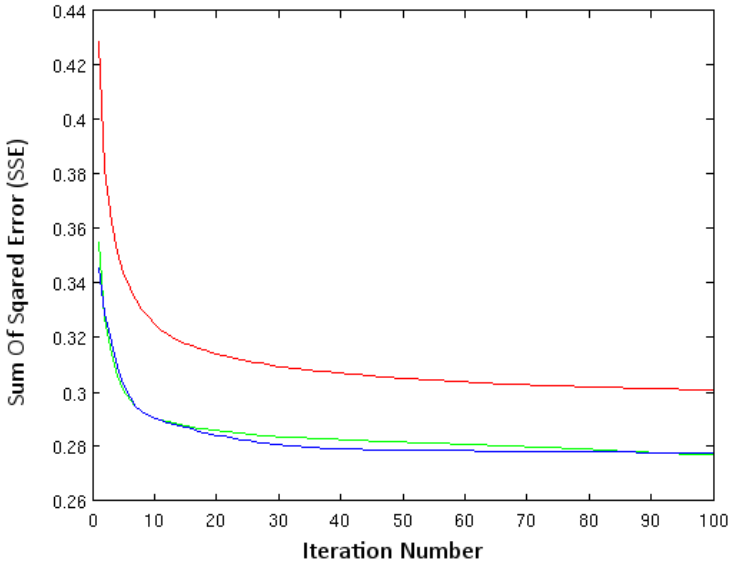


Figure 5.21: Cross test pattern simulation results. Comparison of scale minimised SSE versus iteration number for three different seed phase patterns. (Red) Random, (Blue) Uniform plane phase, (Green) Analytical approximation

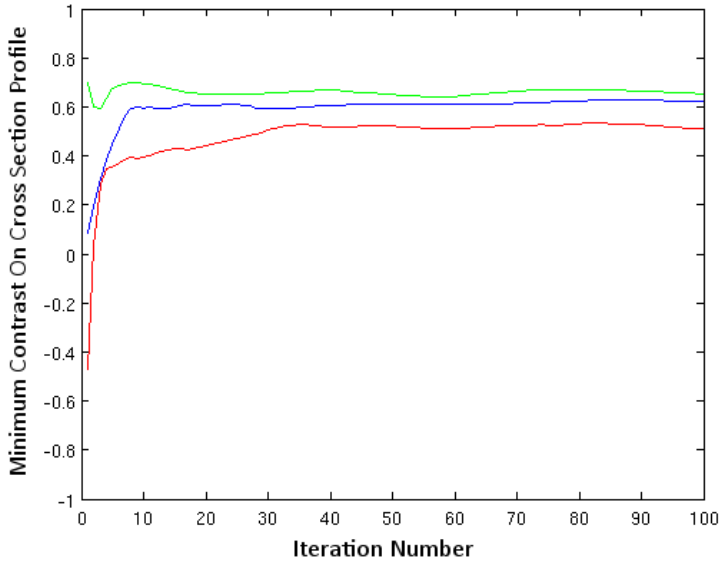


Figure 5.22: Cross test pattern simulation results. Comparison of normalised contrast versus iteration number for three different seed phase patterns. (Red) Random, (Blue) Uniform plane phase, (Green) Analytical approximation

5.6.5 Square Region Test Pattern

This final square test pattern (intensity pattern and holograms in figure 5.23, SSE evolution in figure 5.24) demonstrates the inability of the phase-only constrained planar iterative algorithm to achieve consistent exposure regions for larger non-thin features. Unlike the single line and cross patterns (as well as the bus pattern which will later be shown to improve in section 5.9), the 'speckle-like' variation here is unavoidable in all three cases.

However, the structure of the noise in each of the different seeding cases does vary. The random seed here is shown to have a generally homogeneous noise structure. This useful for active speckle suppression methods discussed in section 5.12. In contrast more structured seeds, such as the uniform phase and zone plate seeds used here, show more structure underlying the noisy pattern generated. In this case the plane phase seed has achieved semi-uniform high contrast regions around the outside of the square image, whilst the zone plate seed has structured the noise into concentric rings.

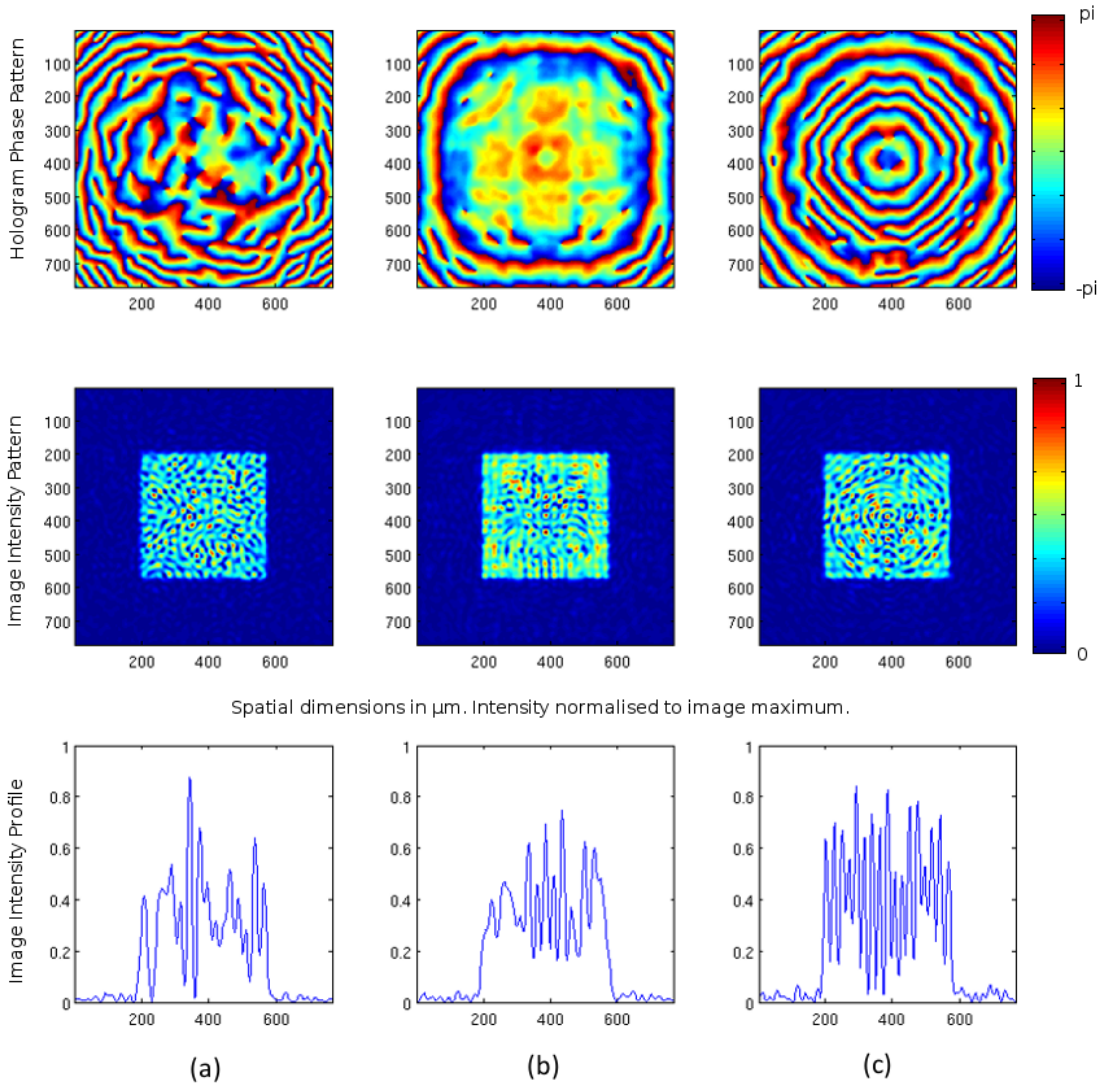


Figure 5.23: Square test pattern simulation results. (Top) Optimised hologram, (Middle) Image intensity pattern, (Bottom) centre profile of intensity, after 100 iterations. Comparison of (a) random, (b) uniform, and (c) analytical approximation, seeded iteration results.

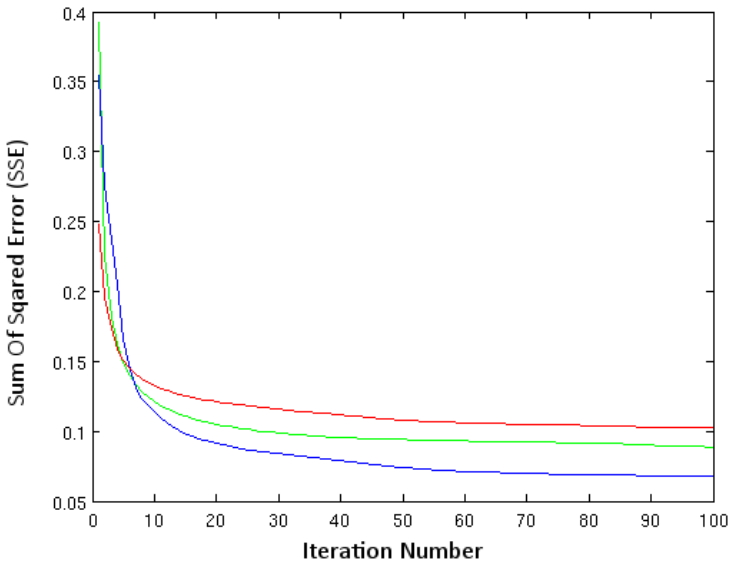


Figure 5.24: Square test pattern simulation results. Comparison of scale minimised SSE vs iteration number for 3 different seed phase patterns. (Red) Random, (Blue) Uniform plane phase, (Green) Analytical approximation

5.6.6 Conclusions From Test Patterns

Structured appropriate seed patterns *can* improve the quality of an image with regards contrast measures. Certain problematic features cannot be optimised away and must be mitigated in other ways. "Speckle-like" patterns are a well known limitation of phase-only holograms [79].

Though more complex seeding structures have shown that in some specific cases more ideal solutions can exist, iteration in these examples improves pattern definition but introduces noise which limits applicability in lithographic situations.

The explanation for the existence of our irreducible 'speckle-like' noise is the inability of the algorithm to optimise phase-vortex [80] patterns out of the hologram and field, resulting in irreducible nulls which severely impact contrast. This explains why random phase patterns - which are much more likely to contain many of these phase vortices - often under-perform compared to the plane and analytical seeded versions in terms of both SSE and contrast.

As mentioned previously, other researchers [56,57,79,81] have experimented with methods to remove these nulls from speckly fields within the confines of a phase-only iterative algorithm. This can be done in several ways. Mixed-Region Amplitude Freedom (MRAF), Offset-MRAF and "Dummy Area" methods all allow for amplitude freedom within certain areas of the field resulting in a less noise-dominated image region. These methods, however, result in large noisy regions surrounding the pattern, making them a poor choice for lithography.

Aagedal et al [79] looks at a basic iterative Fourier transform algorithm (IFTA) with a few different seed patterns and observes similarly that smooth phase functions with no initial phase vortices can help patterns result in noise-free reconstructions. This paper also suggests that pairs of phase vortices can be removed from the field by locating and processing pairs in local regions.

However, our observation here is that a noise-free binary pattern is much easier to produce for these thin binary lines than for larger exposed areas. This result is nonetheless useful as these thin patterns can form useful electronic interconnect patterns as demonstrated in section 7.2. In cases where larger exposures are required we turn to the methods discussed in section 5.12 which allow for noise reduction through signal averaging or reduced coherence, rather than increasing the complexity of the algorithm we apply.

Next, we look to some approaches to extending the usefulness of algorithm and attempt to begin to characterise some limits of the parameters on the hologram design, and the effects of these parameters on image quality and contrast.

5.7 Multi-step Propagations for Simulation Flexibility and Their Effects on Iteration Error

Convolutions are computed by multiplication of the Fourier transform of both transfer function and input distribution. This is performed using FFT operations for computational efficiency. Furthermore, in our simulations we employ GPU processing via the 'Jacket' extension to 'Matlab' on a Tesla C1050 HPC computing card. Described in notation:

$$\mathcal{F}[A(x,y,z)] = a(\nu_x,\nu_y,z) = e^{j2\pi z\sqrt{(1/\lambda)^2+\nu_x^2+\nu_y^2}} \tag{5.4}$$

$$M(x,y) = FFT^{-1}(FFT(L(x,y)) \times a(\nu_x,\nu_y)) \tag{5.5}$$

Where FFT and FFT^{-1} designate applications of the Fast Fourier Transform forward and in reverse. Further to this the image field ($L[x,y]$) embedded in a field of zero values out to the size of $2N$ to ensure a valid Fourier transform from the FFT.

To ensure the validity of the propagation (depending upon window size and propagation distance) it may become necessary to break a longer propagation into several shorter windowed propagations as has been previously discussed by Sypek [20, 82]. Sypek uses the Fresnel approximation to the Rayleigh-Sommerfeld equation and applies its convolution form. We take a similar approach but apply the more accurate Angular Spectrum transfer function and specifically ensure that the AS transfer function will not alias in the sampling regime we have chosen. Since calculation of the transfer function is a one off computing cost its impact on the computation time is low compared to the evaluation of propagations by convolution.

5.7.1 Deriving a maximum propagation distance from the transfer function

Sypek [20] shows how windowing of the field and repeated application Fresnel convolution transfer function can be used to remove aliasing error for longer propagations than can be achieved with a single step. We show here that this can also be applied to the AS transfer function, by considering its maximum propagation distance, derived as follows.

Using the local spatial frequency approximation to the function as defined in [15] on the AS transfer function in the spectral domain:

$$f_l(\nu_x) = \frac{\partial}{\partial x}\frac{\phi(\nu_x)}{2\pi} \tag{5.6}$$

where ϕ is the phase function of the AS transfer function given by

$$\phi(\nu_x) = 2\pi z\sqrt{(1/\lambda)^2 + \nu_x^2} \tag{5.7}$$

therefore

$$f_l(\nu_x) = \frac{z\nu_x}{\sqrt{1/\lambda^2 - \nu_x^2}} \tag{5.8}$$

The sample rate in the the frequency domain as determined by the simulation parameters is

$$\delta\nu = 1/N\delta x \tag{5.9}$$

Where the simulation chosen is N by N pixels and the pitch of the samples is pm. Similarly the extent of the simulated frequency domain is

$$\nu_{max} = 1/\delta x \tag{5.10}$$

Hence the frequency of the transfer function at the edge of the sampled field in the spectral domain is approximated by

$$f_l(\nu_{max}) = f_l(1/\delta x) = \frac{z}{\delta x\sqrt{1/\lambda^2 - 1/\delta x^2}} \tag{5.11}$$

So to satisfy the Nyquist criterion

$$1/f_l \geq 2\delta\nu \tag{5.12}$$

More explicitly

$$\frac{\delta x\sqrt{1/\lambda^2 - 1/\delta x^2}}{z} \geq \frac{2}{N\delta x} \tag{5.13}$$

Therefore the maximum z propagation distance in one step is

$$z \leq \frac{N\delta x^2\sqrt{1/\lambda^2 - 1/\delta x^2}}{2} \tag{5.14}$$

Having propagated this distance we do as [20] and re-window the field to remove energy from outer regions and then repeat to reach longer distances. Ideally the number of steps should be minimised to avoid the inclusion of noise from the windowing function.

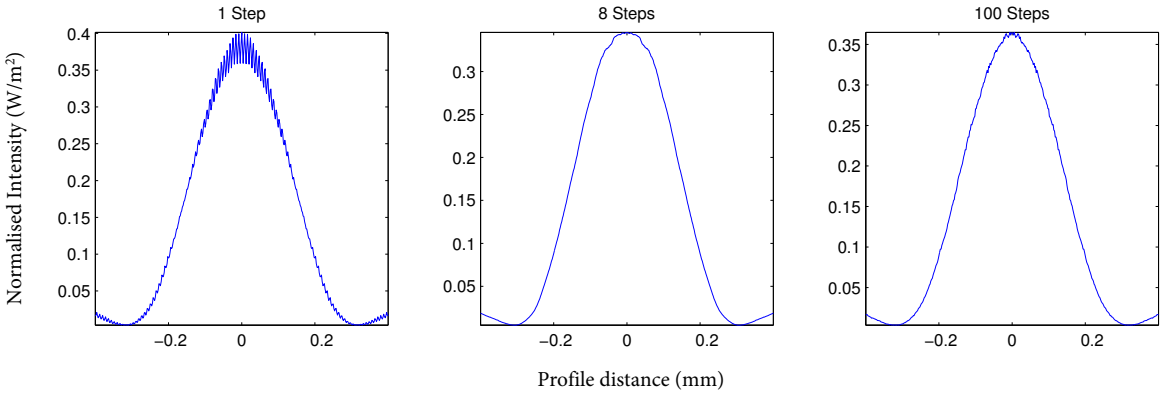


Figure 5.25: Profile of central lobe of intensity diffraction pattern of a rectangle aperture, simulated at 16cm propagation distance. A single step propagation (a) as well as optimal number of steps (8 as determined by the result of equation 5.14) (b) and finally 100 (c) steps are shown.

The practical effect of this modification can be seen in Fig5.25 and it is clear that the erroneous oscillatory noise caused by aliasing of the transfer function is suppressed.

It should be noted that another potential approach is to limit the size of the propagation kernel of the AS transfer function [83]. This potentially improves the computational efficiency of the propagation as repeated steps are not required. However since later in this report we require the breaking of simulations into multiple steps for constraint purposes, the multi-step approach is both simpler and a necessary investigation for application along-side our 3D iteration technique.

5.8 Error Reduction/ Iteration with the AS-TF and Multi-Step

The method of applying a propagation transform kernel in place of an FFT has been applied [37] previously and some [38] have applied Sypek's windowed propagation in iterative algorithms. Discussions of these applications, however, have omitted some of the finer detail in the implementation which can potentially cause problems. Here we firstly examine the differences in error reduction between these methods the basic FFT method and subsequent enhancements which allow us to work a distance closer than the far field, and at reduced computational complexity, specifically examining these assumptions that allow error reduction in the GS algorithm.

Looking at our planar iteration algorithm with the AS transform (Figure 5.26) whilst maintaining a valid sample rate and a single propagation step, we observe a similar error reduction to that of the basic GS method.

For comparison we make one adjustment to broaden the scope of our simulation. By doing a multi-step propagation using the AS transform (as performed similarly by Makowski using Sypek's windowing method), thus enabling longer propagation

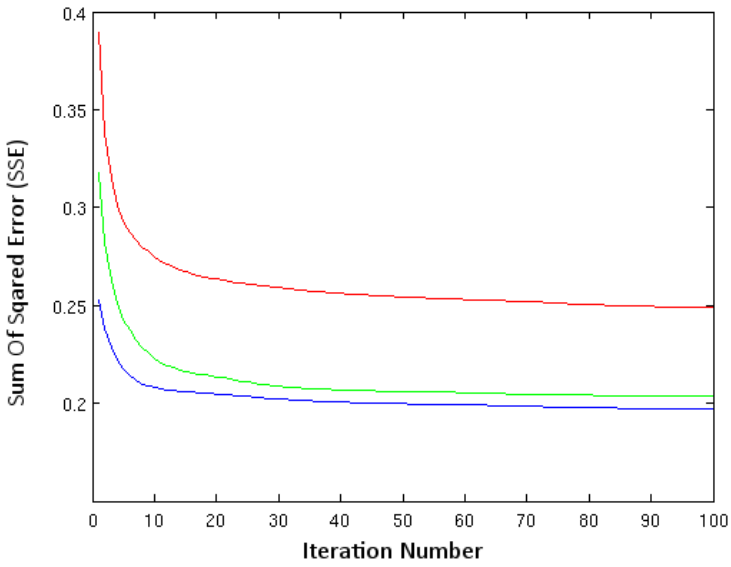


Figure 5.26: Error reduction for single step propagation iteration. Parameters used are identical to those found in the above tests (Table 5.2).

distances with no aliasing of the transfer function. Error against iteration number is shown in figure 5.27.

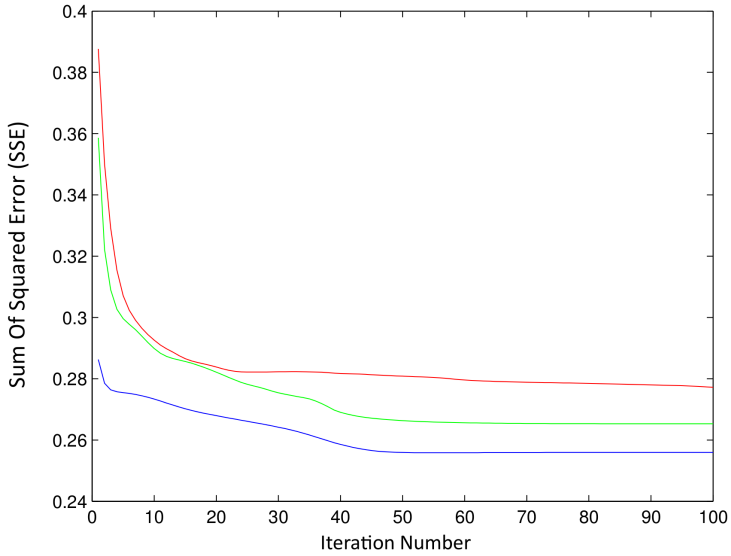


Figure 5.27: Error reduction for three step propagation iteration. Parameters used are identical to those shown in table 5.2 except the propagation distance has been increased to 12cm and the propagation calculation has been broken into 3 equal distance steps.

We will ignore the slight increase in average error from the single to multi-step case, as we would expect error to increase as the size of the PSF has also increased.

Interestingly in this case the error does not decrease as cleanly as the simpler single step case, and in fact in some cases, though hard to see in this example, the error increases from one step to the next.

It is therefore evident that the windowed, stepped, AS transform is breaking some of the rules laid out by the GS algorithm to ensure error reduction. This is a worrying concession for an algorithm which bases itself on the assumption of error reduction.

The constraint being broken here is the complete correction of the amplitude in every constraint plane. Between input and output planes we break the kernel into shorter propagation steps to avoid aliasing the transfer function, and re-window the field whilst letting the amplitude and phase remain the same in the central quadrant. These 'windows' are equivalent to a partial constraint on the amplitude (setting it to zero) and this can cause a change of information outside of the measured constraint planes which results in the apparent increase of error in the image i.e. constraints are applied to the field which do not work to the advantage of the SSE decrease on the hologram or image planes.

Thankfully, despite this loss, all of our tested examples converge, with only a small upward variation in error.

Finally, we compare here a valid sampled one-step propagation with a two-step counterpart, which has been broken into 2 and 10 simulation steps of 1/2 and 1/10th the propagation distance respectively. Ideally this should result in an identical error reduction, and indeed figure 5.28 shows only a very small variation in the pattern of error reductions between the single and multi-step simulations.

What we have therefore shown here is that the methods of [20] and [38] can be applied with the AS transfer function in place of the Fresnel convolution propagation, with minimal error introduced. The convergence of the iterative algorithm is affected but not to such a degree to prevent the algorithm from reaching a useful or accurate solution.

5.8.1 Filtering and Pixelation as constraints in Iterative Optimisation

Expanding upon basic analytical simulation sampling criteria and turning to more complex iterative design processes, we again look at the case wherein we expect the optical bandwidth to be higher than the simulation bandwidth, determined by the chosen sample pitch. Doing so it becomes apparent that the iteration result is far from accurate when compared to the physical experimental output. Figures 5.29 (a) and (b) show such a case where optical bandwidth is higher than simulation

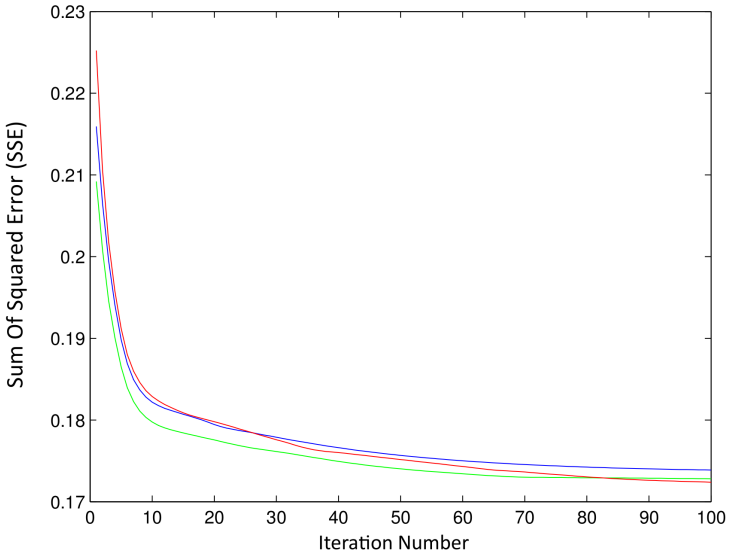


Figure 5.28: Error reduction where propagations have been broken into 1 (green) , 2 (blue) and 10 (red) steps. Parameters used are otherwise the same as those shown in table 5.2.

bandwidth and iteration using these parameters results in a very attractive looking profile with a high contrast and low noise levels.

Practically, however, this is far from the truth. Figure 5.29 (c) and (d) show that by increasing the resolution of our simulation to a suitable rate, a variable noise is made evident.

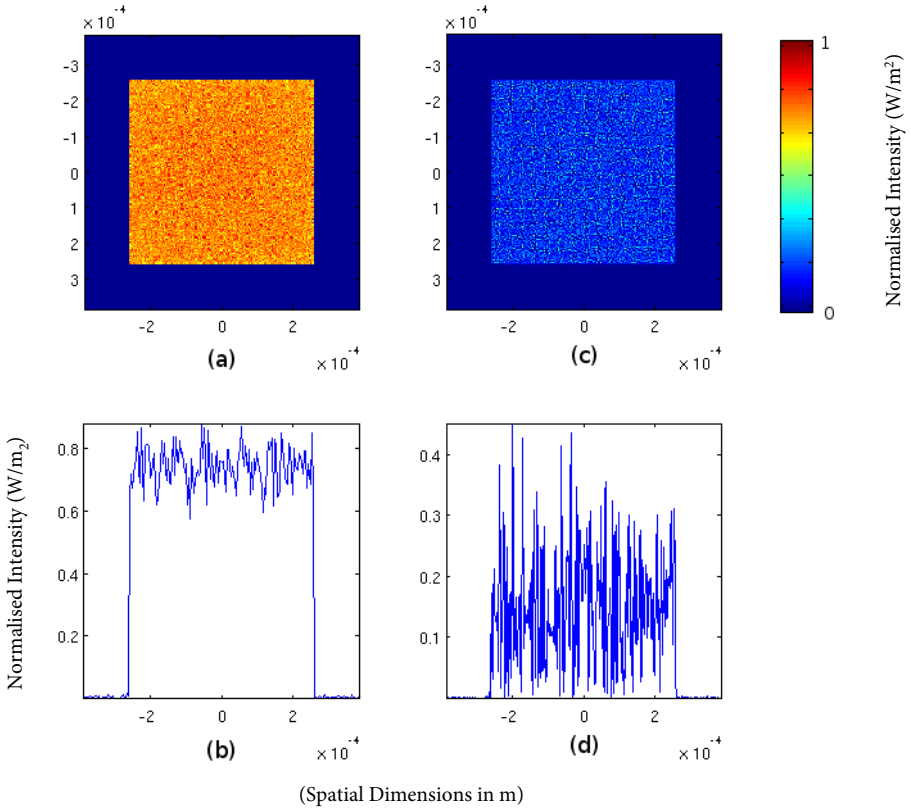


Figure 5.29: (a) Image and (b) profile of under-sampled iterated hologram image in simulation. (c) Image and (d) profile of simulated reconstruction at a more than sufficient sample rate.

When the optical resolution exceeds the simulation resolution, the subsequent optimisation is based on an incomplete description of the field. As a result the apparently 'optimised' field is far more attractive than is in fact delivered by the actual hologram. The question then becomes: How can we ensure that this is not the case, and that the optical resolution does not ever exceed the simulation resolution? This would leave our simulation (and thus optimisation) as correct as possible.

Another, more practical, constraint must also be taken into account. If we have a known modulator bandwidth which limits the bandwidth of the system to a rate lower than the optical bandwidth, then there is no reason to simulate at a sample pitch higher than this known rate, as this would expend computational effort where not necessary. However, we know that if the simulation sample bandwidth does not exceed the optical bandwidth the results become unrealistic. The result of this is that we must consider which is lower - the optical or modulation bandwidth - and sample our image appropriately.

As an example for a square hologram with 8192μm side length, we may choose

a minimum focal distance of 8cm therefore by the equations for resolution in 5.1:

$$\delta x = \frac{405 \times 10^{-9}}{2 \times 1 \times \sin(\tan^{-1}(4096 \times 10^{-6}/8 \times 10^{-2}))} \approx 3.96\mu m \tag{5.15}$$

In the simple case we would need to have a simulation sample pitch of at most, $3.96\mu m$ to correctly represent the image intensity pattern. However iterating a field at this pitch and then taking the square of the magnitude of the field to simulate the intensity again results in an inaccurate pattern, because scalloping of the signal with a sample rate lower than 0.5 times the bandwidth of the optical system will hide noise present in the real experimental evaluation of the image. This in turn is due to the simulation having no limit on spatial frequencies as part of the optimisation.

To overcome this problem such that we can move the image plane close to the modulator (.i.e have the optical bandwidth not be the limiting factor), we have determined that we can limit the bandwidth of the hologram to more closely match the limitations of the physical device within the iteration scheme.

Two approaches to doing this have been considered, low pass filtering of the field in the spectral domain, and pixelation of the hologram in the space domain.

5.8.2 Filter

The first of these modifications is to use a filter to limit the bandwidth of the field simulation to at most half of the bandwidth of the overall simulation. Fig 5.30 (c) shows our AS transfer function, this time filtered by a circular bandlimit filter that blocks any frequency above 0.5 times the bandwidth of the simulation. This is applied as a limit to the bandwidth of our propagation kernel, ensuring that angles of plane waves above this bandlimit are ignored and removed from each propagation.

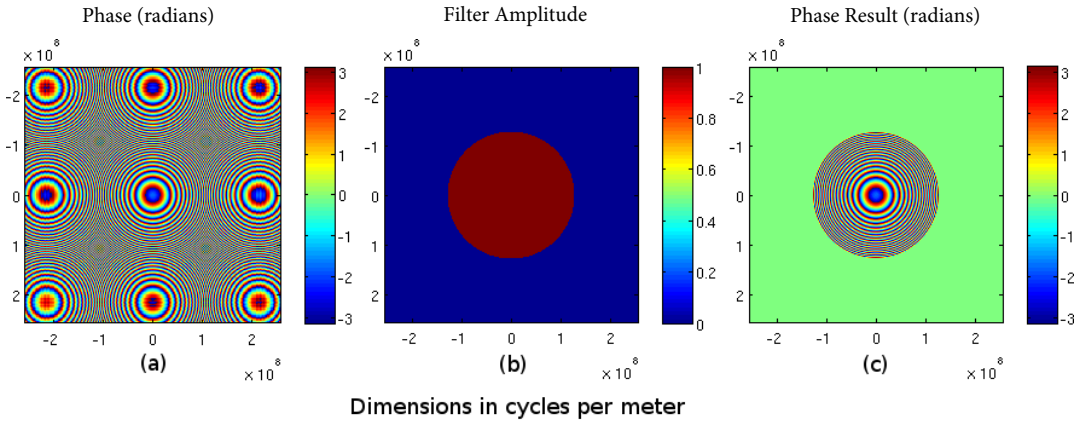


Figure 5.30: Example AS propagation kernel ($A(x, y, z)$), (a) Phase distribution, naive approach, (b) circular filter, (c) filtered kernel function

5.8.3 Pixelation

In addition to this filter constraint we can also apply a pixelation algorithm in cases where the simulation sample pitch exceed the modulator pitch, allowing a more accurate model of the implemented SLM to be simulated, as long as the ratio between the two sample pitches is some integer value. To do this we apply a function which for each SLM pixel (which may be several simulation pixels in size) re-samples the field and sets all the simulation pixels within the region of that SLM pixel to the same value. This constraint is applied to the hologram only when it is modified in each iteration loop, prior to the phase-only constraint. This is shown diagrammatically in Fig 5.31. This modification ensures that the field propagated away from the modulator is accurately implementable on the SLM, and more accurately models the field produced by the SLM.

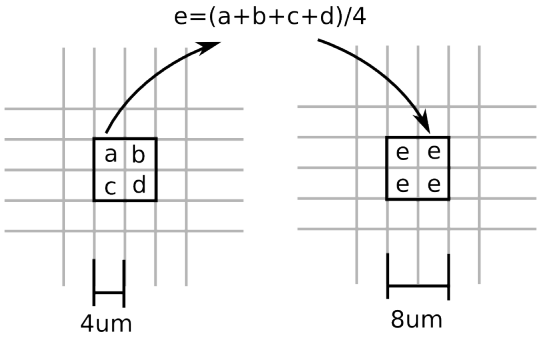


Figure 5.31: Pixelation routine applied to every group of 4 $4\mu m$ pixels to produce $8\mu m$ squares which match the $8\mu m$ SLM pixel size

5.8.4 Analysis of Filtering and Pixelation Modifications

Both of these approaches allow us to more closely model the re-sampling function that must occur to implement a hologram where the field bandwidth is limited by the modulator and not the optical system (i.e. the focal distance). Importantly they also allow us to simulate a limit on the bandwidth of the device without foregoing accuracy in our simulation, or requiring us to post process the iterated hologram to a higher sample rate.

For our comparison of these methods, the implemented hologram has a size of 1024×1024 SLM pixels at $8\mu m$ pitch. This is sampled at $4\mu m$ in simulation pixels. At the sample pitch of the simulation here the resulting intensity pattern is under-sampled. Iterative algorithms are applied with no filtering, a filtered transfer function, and both a filtered transfer function and a pixelated hologram constraint. A fixed 100 iterations are used in each case. The iterative algorithms are seeded

with the same random phase distribution each time. After the generation of each hologram, two analysis simulations are performed; one at the sample pitch of the simulation, and an oversampled simulation at $2\mu m$ sample pitch where the square $8\mu m$ sampled version of the hologram is interpolated to $2\mu m$ by nearest neighbour interpolation to model the experimental implementation on an SLM.

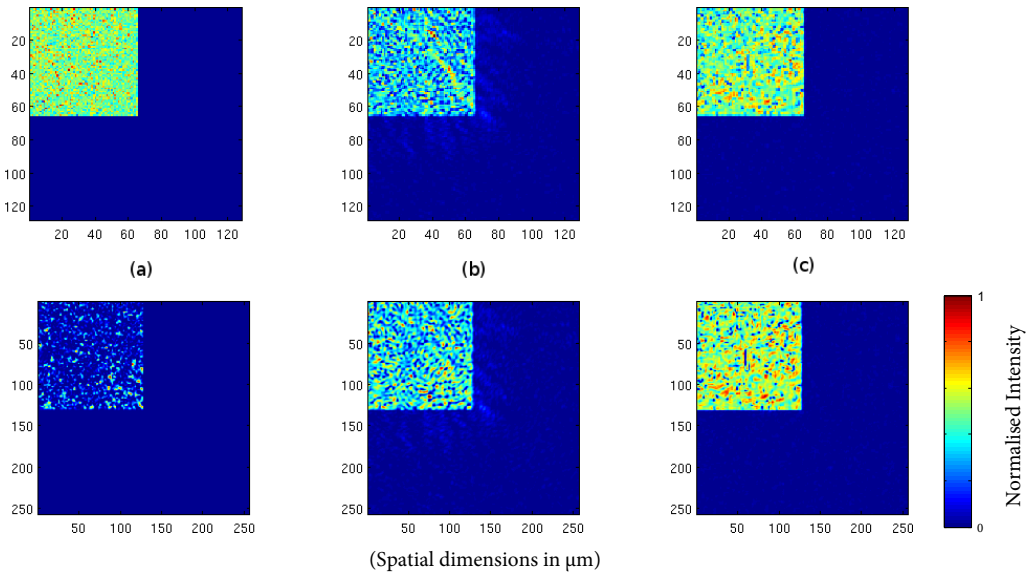


Figure 5.32: Comparison of simulated image at iterated resolution (Top), and higher resolution analysis (Bottom) for iterated holograms generated by (a) naive approach, (b) filtered iteration, (c) pixelated iteration.

Comparing the lower and higher resolution image pairs in Fig5.32 we firstly note that in the naive approach (a) we observe, as above, a large inconsistency between simulation at the iteration sample pitch and its sufficiently re-sampled analysis. Secondly, we note that this is not the case for modified methods (b) and (c), and that even the smallest features shown in the higher resolution analysis are also present in the simulation at the iteration sample rate. This is important as next we note that the quality of the image improves significantly in (b) and (c) compared to (a). This is because the additional constraints on the iterative process allow for a more realistic simulation given a limited bandwidth, whilst also limiting the computational scope of the algorithm, and more realistically representing the implemented hologram. In addition by inspection the filtered and pixelated case shows a more uniform output, though vortices still remain.

Table 5.3 shows a comparison of the coefficient of variance (COV) for each pattern in its intended exposed region. This is applied as a quality metric as the smaller the variation, the better the intensity pattern. The COV for the naive approach is the largest, with the filtered approach improving significantly and the pixelated approach does indeed perform best with regards this measure.

	Naive	Filter	Pixelate
CV	0.7655	0.4397	0.3430

Table 5.3: Normalised standard deviation (coefficient of variance) calculated for the square test pattern in the intended exposed region for each method in the high resolution simulation

5.9 Hologram and Image Size

In addition to limitations in bandwidth affecting the final solution, the size of the hologram - regardless of bandwidth - also has a serious effect both on the kind of patterns which can be achieved and also the convergence of the GS algorithm to a useful solution. To examine this issue we take a situation much like that of the multi-line bus tested in section 5.6.3. As we have shown previously, this is unable to form a contiguous pattern under the size constraints used in that experiment. We simulated a set of bus patterns with increasing hologram size allowing more distant hologram fringes to form. In this case a plane phase is used to seed the hologram iteration.

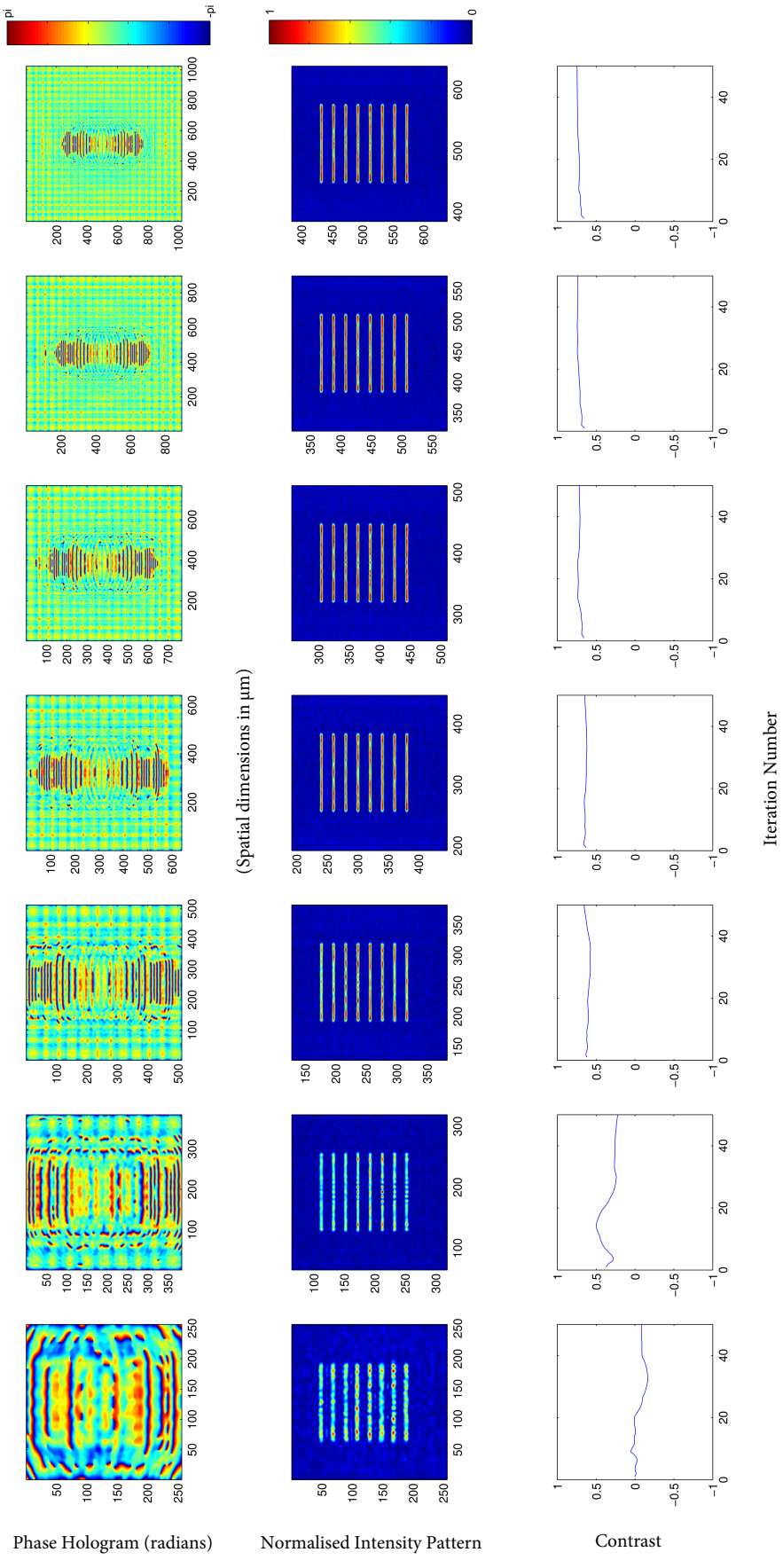


Figure 5.33: (Top) Hologram. (Mid) Simulated intensity image. (Bottom) Pattern contrast over iteration number. For an iterated hologram with hologram size varying from $768\mu m$ (left) up to $3072\mu m$ (right).

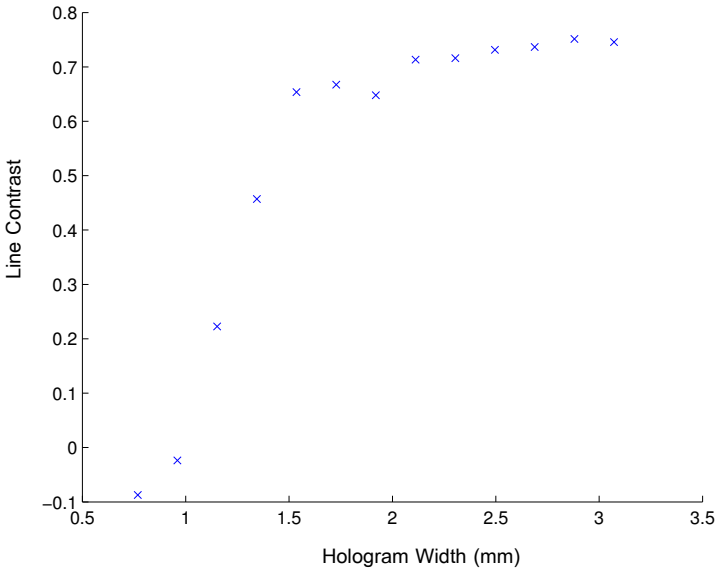


Figure 5.34: Contrast for a range of hologram sizes for a small dense bus pattern.

Figure 5.33 shows such a progression in hologram size and figure 5.34 shows a summary of the final contrasts arrived at after 50 iterations for the same range of hologram sizes. As we increase the width of the hologram (whilst maintaining the bandwidth of the system with a filter as applied in section 5.8.1) we note the increase in minimum and average contrast i.e. the improvement of the image quality from below zero to up to approximately 0.75.

It is interesting to note that under this filtered approach, the plane phase seed case shows a localisation of fringes to the transverse vicinity of the image pattern. This localisation is explained by considering that by filtering the hologram we limit the maximum diffraction angle that can be achieved to move energy from one part of the transverse hologram field, to another region in the image/object plane, i.e. the shape of the PSF of the system at some distance from the focal plane.

If we were to relax the filter to allow higher frequencies, higher frequency regions will appear at the outside of this current localised hologram region and diffract more of the outlying energy into the image pattern, which would also result in a smaller or more well defined feature from an effective increase in numerical aperture.

Essentially what we are observing here is a transition from an optical (NA size) bandwidth limit to a synthetic filtered limit applied in the iteration propagation, but interestingly this filtered limit has localised our image pattern and hologram pattern, and allows for significant improvement in the pattern quality, through increasing the size of the hologram.

Furthermore it is interesting to note that increasing the size of the hologram

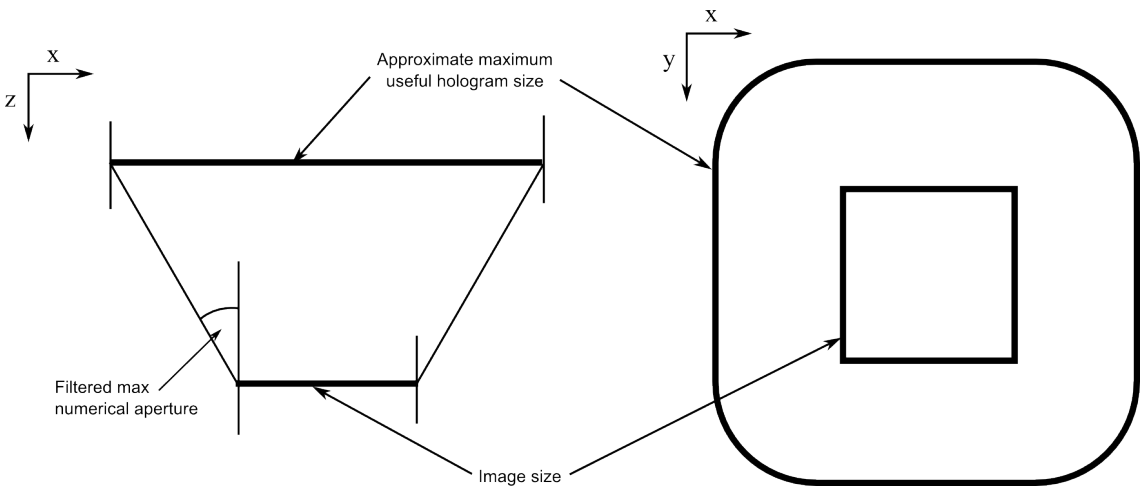


Figure 5.35: Geometrical construction of the loci of points which can be reached by energy from a filtered angular spectrum propagation from image to hologram, giving the maximum useful hologram size. (Assuming a completely coherent field)

beyond allowing the full localised hologram to develop has little/no effect on the image quality whatsoever as shown by the largest holograms shown in figure 5.33. This leads to some interesting potential conclusions about how for a fixed maximum bandwidth hologram, the minimum size of hologram which will not impact the image quality can potentially be calculated by considering a loci around image points given a maximum diffraction angle. A diagram demonstrating a potential optical constraint can be viewed in figure 5.35.

Mathematically we approximate the size of the required hologram

$$\begin{aligned} W &= I_W + 2\delta_W \\ \delta_W &= z\tan(\theta) \\ \theta &= \sin^{-1}\left(\frac{\delta x \lambda}{p}\right) \end{aligned} \tag{5.16}$$

Where W is the approximate maximum useful hologram width (assuming a square hologram), I_W is the image width, δ_W is the increase in hologram size around the edge of the image, z is the propagation depth from hologram to image. θ is the effective aperture angle derived from the effective numerical aperture which we will define in section 6.3.4.

For the hologram parameters used in this simulation, the above equation we approximate a useful size of 1.8mm which is larger than that of the point of change in gradient shown in figure 5.34.

5.10 Image Feature Width

Though we observe "speckle-like" noise when attempting to create features that are much larger than the PSF of the system, as shown above it is possible to produce thin continuous features with controlled characteristics when applying an analytical or plane phase seed seed to a thin image pattern. The concept of 'thin,' however, is ill-defined and so here we investigate the limits of image feature width with respect to noise and contrast.

A series of iterated holograms were produced with a varying binary image line width, and the simulated line cross sections recorded and analysed. The results are shown in figure 5.36.

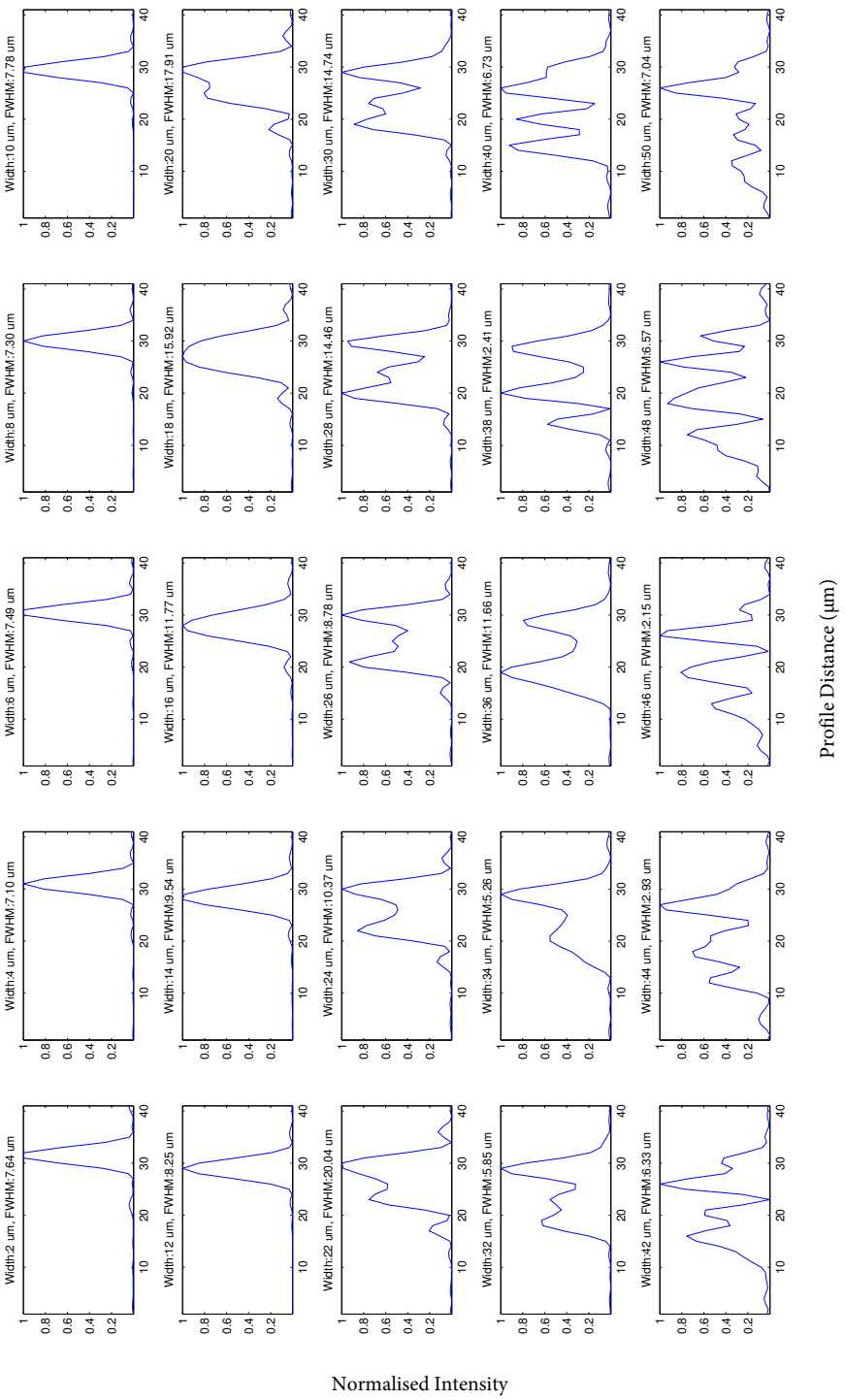


Figure 5.36: Graphs of simulated cross-sections of features with increasing intended feature. FWHM measured width is shown for comparison.

Our standard contrast measure is difficult to calculate for this kind of profile since the choice of feature width can wholly determine the outcome of the parameter for this Gaussian shape because it is impossible to ignore the edges of the pattern on such a small scale. Instead then, we look at a threshold binarised pattern to determine feature width, and choose the 'full width at half maximum' criterion to do so. Where multiple peaks protrude above and below the half-height width, it can reasonably safely be assumed that the pattern is far from ideal for lithography and so the width of these patterns becomes unimportant.

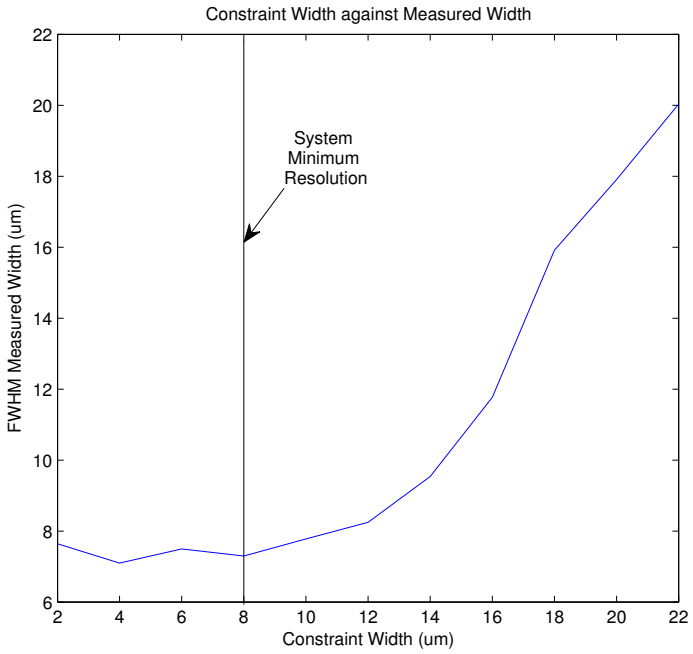


Figure 5.37: Summary graph of varying FWHM line width with intended constraint line width.

From figure 5.37 it is clear that adjusting the width below the PSF of the optical system ($8\mu m$ in this example) has little effect on the final pattern. However thin the intended image is below this threshold, the optical system cannot achieve that level of feature resolution. However, it does not degrade the image profile. Above the width of the PSF of the optical system, we have some control over how wide the image may become, but this is severely limited and unpredictable when optimised with the iterative algorithm. In this case this is observed at line widths of above $20\mu m$ or 2.5 times the width of the PSF. This confirms and quantifies our previous assertion that thin line based patterns are possible without the need for speckle removal systems.

5.11 Image Background Value

When creating an ideal constraint image the most simple approach is to create a binary image with the values '1' and '0'. However, as briefly touched upon in the multi-planar experiments of [38], the effect of the chosen background value has a serious effect on the result of the algorithm in 3D. We will come to discuss this in section 6.3.1. Though this effect is very serious in the multi-plane case (owing to potential loss of information), it is also potentially easy to miss that - for the thin geometries just discussed - it also has a serious effect on the general image properties and to some extent can mitigate some speckle type noise even in the 2D case. We proceed to show here that adjusting the background value not only affects the general background to foreground intensity ratio, but also has a significant effect on the ability of the iterative algorithm to converge to a noise free solution.

To examine this effect we consider our simple line and bus test patterns with feature width approximately that of the PSF of the optical system. A series of profiles is then computed with varying background value. The image "high" value is set to amplitude 1 and background "low" value to amplitudes between 0.0 and 0.3.

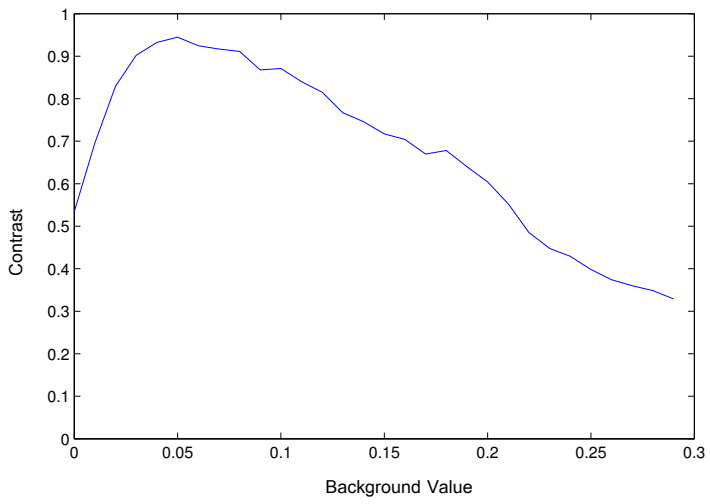


Figure 5.38: Line contrast parameter varying against varying background value.

Figure 5.38 shows how the contrast for a line bus test pattern varies for a range of background values. From this graph we note two specific regions. Firstly, the sharp increased in contrast value as the background value initially increases from zero. Secondly, the slower decrease in value after a peak at approximately 0.05 background value. In the ideal case this value would merely decrease from a maximum value of approximately 0.98 at 0 background value.

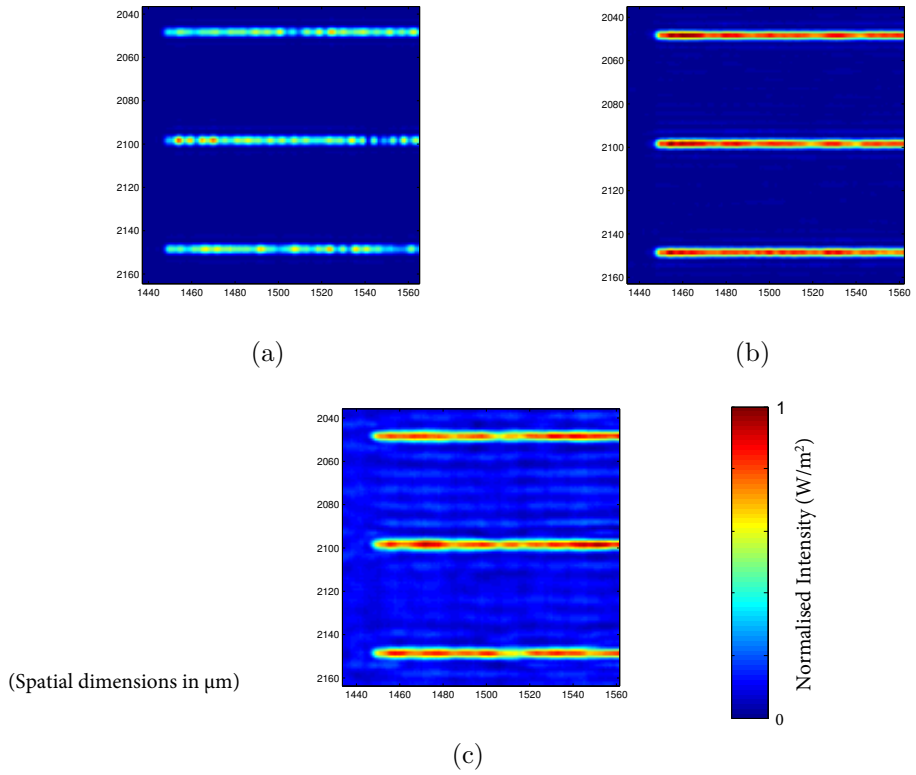


Figure 5.39: Simulated examples of iterated line intensity patterns with background values of (a) 0.0 (b) 0.05 and (c) 0.2

Figure 5.39 shows three examples of low, optimal and high background values. Note that at low (0) background we observe not an increase in background noise but more of the familiar noise. This is not the case for the other examples wherein we observe a well defined contrast. For the yet higher background value we observe the expected decrease in background/foreground ratio.

5.12 Noise suppression for phase-only holograms

5.12.1 Noise reduction

Thus far we have shown how thin features can be achieved and how varying numerous parameters can help us to optimise a pattern for lithographic exposure. For larger features, however, we have seen noise inherent to phase-only hologram construction. We now need to find a way to suppress this problematic intensity variation, thus allowing us to expand on our method. Various optimisations exist for the removal of speckle noise in coherent optical projection. The noise we observe here is not the traditional optical definition of speckle, caused by coherent light scattered by a rough surface, but it does follow a very similar structure [15]. Therefore some traditional speckle suppression mechanisms can be applied.

Some examples of methods are vibrating screens, extended sources and rotating diffusers i.e. methods that affect the effective coherence of the system. One much simpler approach in our case, however, has been to use the active electronic nature of our projection to average multiple noisy patterns over the course of the exposure. This has been previously demonstrated by [53, 54] in 2D using a far-field FFT approximation. Here we apply it with a lens-less holographic projection giving a more flexible method.

5.12.2 Time division multiplexing

Because our phase modulator is a computer controlled display which updates 60 times per second, we can display a number of different holograms over the course of an exposure. The photo-resist absorbs an energy density pattern which is the average of the intensity patterns applied. This process gives us an improvement in noise by applying effective signal averaging. For some optical systems we have constructed, it is possible to require exposure times of up to 20s because of the amount of energy lost by filters (spatial and energy) and all of the reflective surfaces in the optical system. This long exposure time is not ideal but could easily be reduced in a more practical system by increasing the source power.

In order to generate a number of unique speckle patterns we ran our iterative process multiple times with a different random seed phase pattern in each case. The noise observed here is deterministic, and depends on the structure of the hologram. It is therefore different for every seed pattern.

A simulated experiment was performed in which a square pattern was iteratively optimised using a planar multi-step optimisation algorithm. 50 holograms were generated and the minimum contrast was measured after each intensity pattern is

Propagation distance (z)	Wavelength (λ)	Sample/Pixel Pitch
$8cm$	$405nm$	$2\mu m$
Image Width/Height	**Simulation Width/Height**	**Feature Width**
2048 pixels	4096 pixels	$256\mu m$
Constraint Low Value	**Constraint High Value**	**Number of Iterations**
0 to 0.2	1.0	50

Table 5.4: Parameters used in planar iteration experiments for testing signal averaging

added. The parameters of the system and simulation are summarised in table 5.4. Background value was also varied and contrast output was averaged over 10 trials.

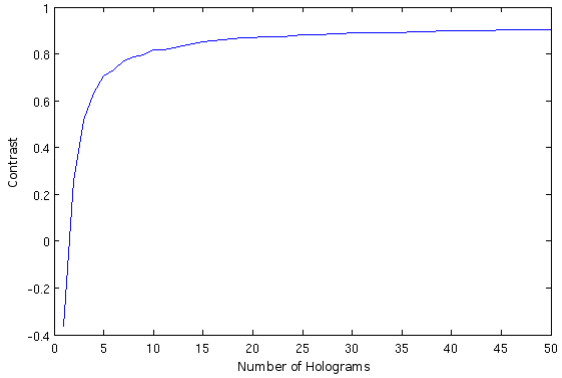


Figure 5.40: Number of Holograms vs Contrast for a fixed example image averaged over 10 trials.

Fig 5.40 shows a graph of "number of holograms" vs "contrast". This has been averaged over 10 trials to suppress some of the random variation associated with choosing a random seed. The simulation sample pitch used was $2\mu m$ and the "filter" method was used to limit feature size to a minimum of $4\mu m$ though optical resolution is limited to a yet larger size of $7.9\mu m$. This approach is able to far improve our image quality with regards this metric. Furthermore, as opposed to a partially coherent method, no resolution is lost.

Figure 5.41 shows final attained contrast against background ratio value for this simulation set-up. It can be seen from figure 5.41 that the contrast obtained using different background values displays a similar potentially counter-intuitive trend when the background value is close to zero (i.e. when the intended contrast is this high the attained contrast is low) as our other background experiment. This demonstrates that this zeroing of low region energy - and thus limiting of degrees of freedom of the pattern - also has an effect on the average contrast and not just the

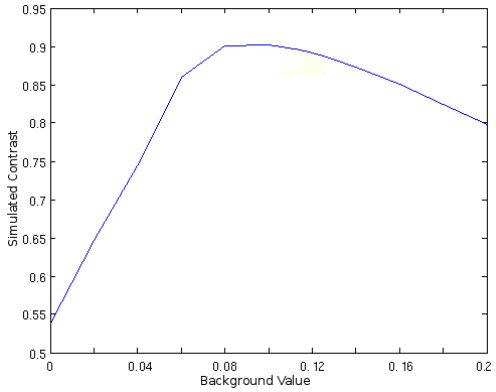


Figure 5.41: Background image value versus contrast using 50 holograms averaged over 10 trials

convergence in the case of the thin features. Here we see that the pattern reaches a maximum intensity at a background value of 0.1.

5.13 Coherence Control

Control of coherence is a widely adopted method of coherent speckle suppression [80]. Lithographic projection (and more recently proximity [58]) exposure systems will optimise source patterns to control spatial coherence. This allows for extra parameters of adjustment to optimise masks for dense or sparse features, or particular types of geometry (Manhattan etc). Some simulations and basic experiments we have carried out provide proof of concept for these methods being usefully integrated into 3D lithographic methods. In general, spatial coherence is adjusted by changing the shape of an effectively incoherent source. This can be done either by limiting the shape of a completely incoherent source, or by using a diffuser to randomly vary the phase of a coherent source.

To simulate the effect of an incoherent source a number of different illumination patterns equating to varying angles of illumination can be applied to a hologram. Simulating propagation to the focal plane for each of these patterns and then summing the intensity pattern for each (i.e. an incoherent sum) gives the effective overall intensity pattern if it can be assumed that no coherent interference occurs between each effective source point.

Figure 5.42 shows the image produced by a single iterated hologram, compared with the same pattern incoherently simulated by summing the output pattern for a range of sampled input source points. In this case the important system parameters are a square hologram width of $2048\mu m$, with a distance of $8cm$ from hologram to focal image and a $10cm$ distance from source to collimating lens. The image is a

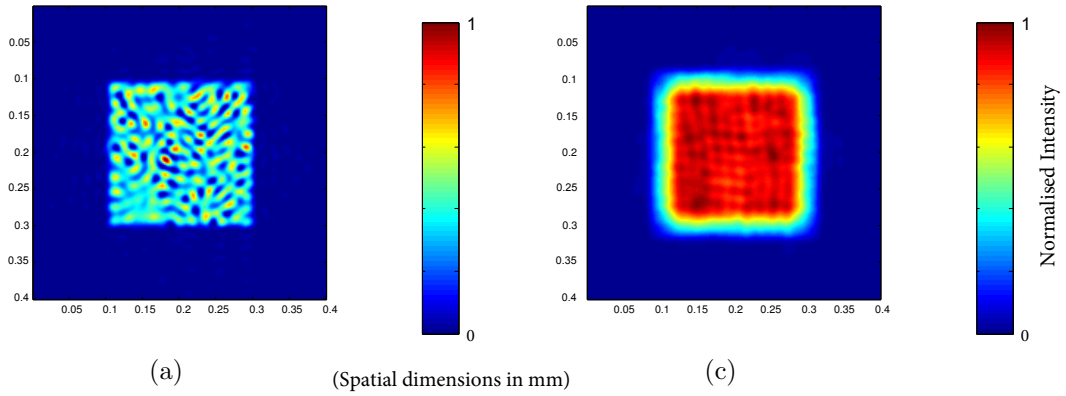


Figure 5.42: (a) Fully coherent and (b) partially coherent test image intensity simulations.

$200\mu m$ width square. The source is simulated as a grid of 10 by 10 points with a total size of $50\mu m$. It is clear that a trade-off is occurring here: As we increase the size of the source the contrast increases but the resolution of the system and its edge definition decreases.

5.14 Noise Reduction Conclusions

The partially coherent method for speckle reduction has proven to be less useful for our test lithographic system thus far. It requires the inclusion of either an incoherent source or a rotating diffuser. In addition, using a partially coherent method also reduces system resolution. On the other hand, the active signal averaging method requires no modification from our existing active projection set up and does not reduce resolution.

Despite this, in a real mask aligner situation (where an incoherent source is used to increase in coherence to a suitable level) adjustment of the source shape or system coherence is likely an appropriate approach to gain 3D modulation and resolution at a distance from the mask. This must be weighed against a potential increase in noise.

5.15 Simulation and Experimental Comparison

A hologram for a bus pattern of eight lines with a pitch of $32\mu m$ has been generated to show correlation between simulated and experimental results. The parameters of the simulation are shown in table 5.5. The optical bandwidth of the system described by these parameters predicts a feature width of $7.91\mu m$. The transfer

Propagation distance (z)	Wavelength (λ)	Sample/Pixel Pitch
$8cm$	$405nm$	$2\mu m$
Image Width/Height	**Simulation Width/Height**	**Feature Width**
2048 pixels	4096 pixels	1 pixels
Constraint Low Value	**Constraint High Value**	**Number of Iterations**
0.025	1.0	50

Table 5.5: Parameters used in planar iteration experiments experimental verification.

function of the AS propagation has however been filtered to allow a minimum of $8\mu m$ features. This then matches with the $8\mu m$ pitch of the 'HOLOEYE' Pluto SLM used to project the image.

The phase hologram generated by this iterative setup is shown in figure 5.43. Figure 5.44 shows subsections of the bus pattern in simulation and experiment and figure 5.45 shows profiles of one of the lines of the bus pattern. The profiles and intensity images make it clear that the pattern high and low regions are largely maintained when implementing the hologram in experiment. The measured contrast in the experimental image profile is 0.65 where as in the simulation profile we observe a contrast of 0.89. This reduction in contrast appears to be because of a relative increase in background noise level rather than an increase in high region noise variation. This is to be expected as in experiment, non-ideal beam profile, less than 100% SLM fill factor, reflections and non-ideal modulation are all likely to increase background noise. Despite this, the contrast is still more than sufficient to produce a well-defined lithographic exposure.

Interestingly we observe a slight change in general profile shape from simulation to experiment. The blue simulated profile bows in the central region of the line with an increase in intensity towards the ends of the line of approximately 0.2 on this normalised scale. The experimental profile does not show this same distortion and remains at a relatively flat intensity over its length. It is likely that this is caused by the Gaussian profile illuminating the SLM from the spatially-filtered laser. In this case the pattern would work to our favour, reducing some intensity at the edges of the pattern.

Measured FWHM feature size of the experimental pattern is $7.85\mu m$, which matches relatively closely with the simulated pattern showing approximately $7.32\mu m$ feature size (though both vary slightly along the length of each line).

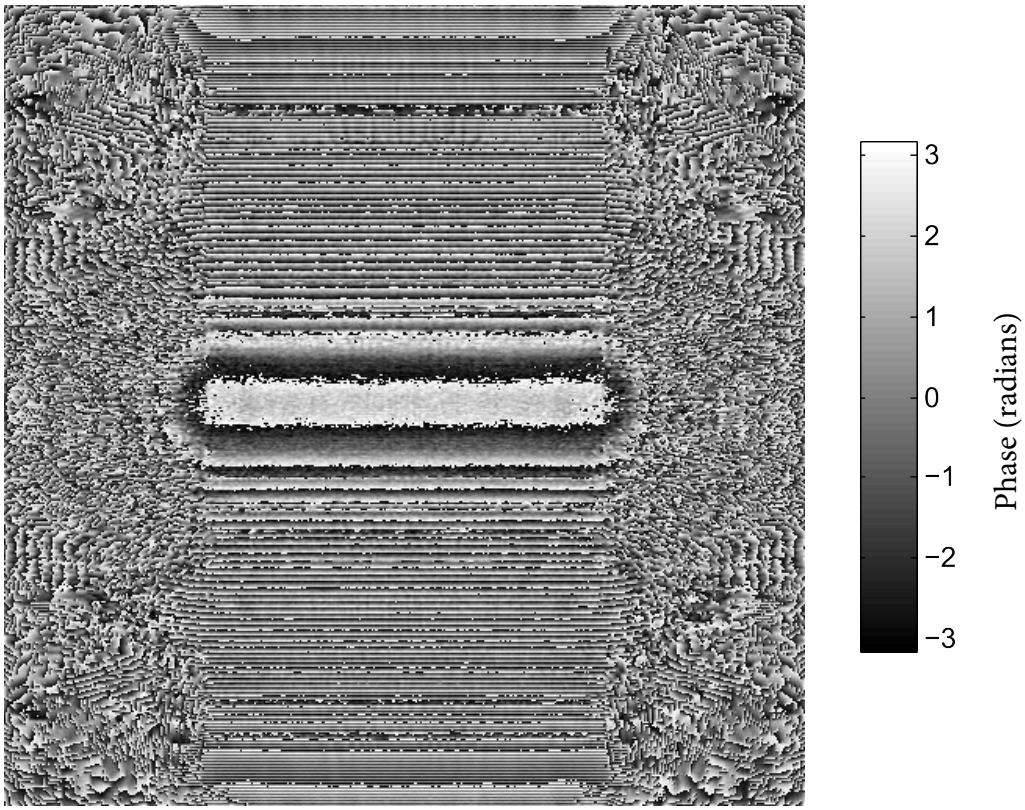


Figure 5.43: Bus lines pattern hologram phase.

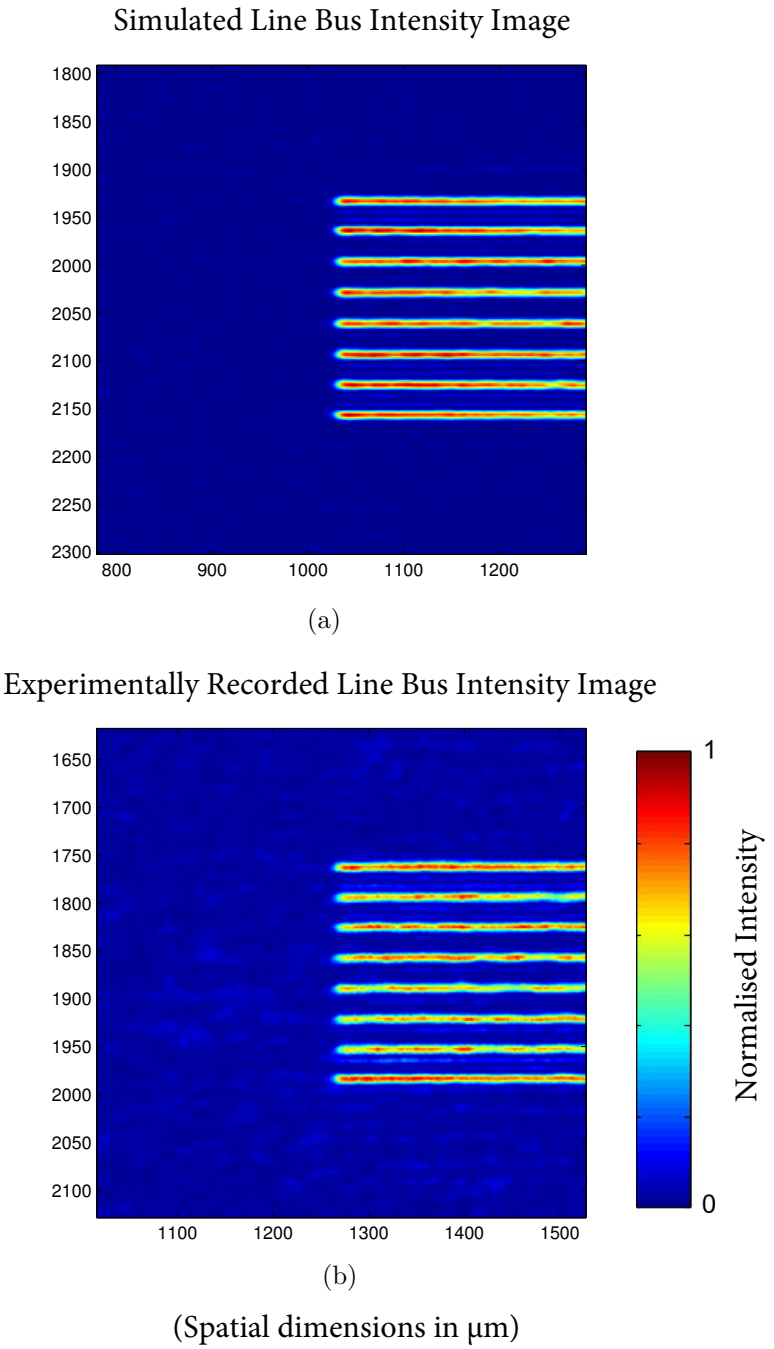


Figure 5.44: (a) Simulation and (b) experimental reconstructions of bus hologram image intensity.

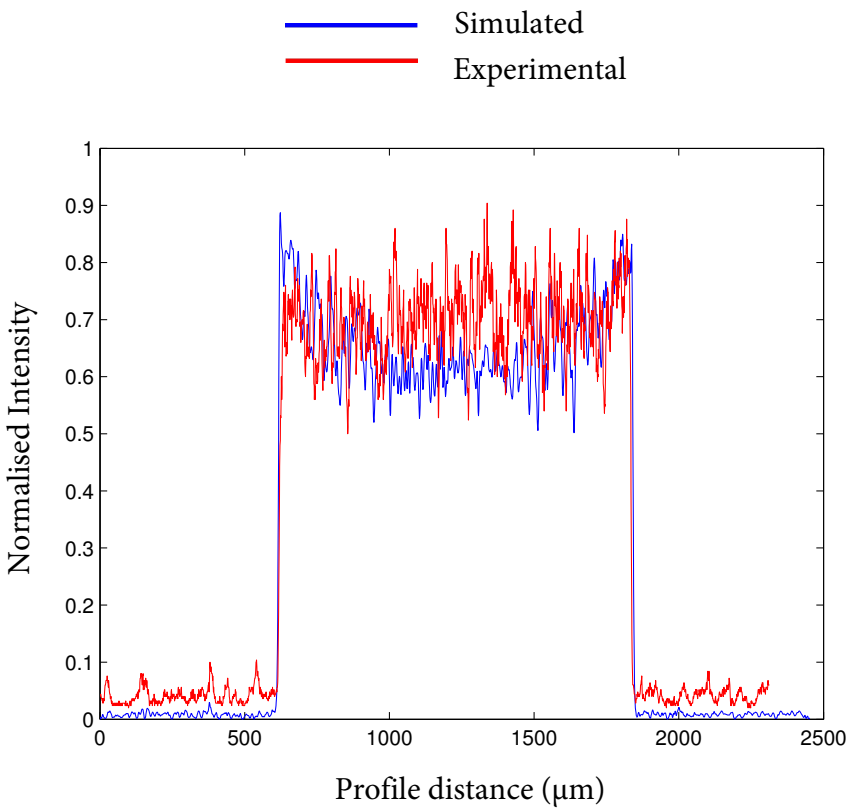


Figure 5.45: Profiles of a single line from simulated and experimentally created bus images.

5.16 Conclusion

In this chapter we have firstly formalised and simulated critical limits on image resolution and therefore simulation resolution. Furthermore we have shown how simple filtering applied to the transfer function kernel can stabilise feature size below a practical implementation limit. We have then gone on to define our metrics of quality, namely SSE and minimum contrast (alongside some more standard measures of variation). We applied these to the analysis of both the iterative optimisation process and the images produced by such a process.

We have outlined both a basic lensed GS iterative scheme and discussed its pitfalls, such as pattern noise and undiffracted noise collection. We have then gone on to define a lensless iterative scheme using the AS transfer function kernel. Following this we set up a collection of different patterns discussing the features, advantages and disadvantages of such iterated patterns, and compared them to analytical patterns of the same scope and scale. We have been able to show how the iterative method when set up and initiated correctly is able to produce well-defined and delimited patterns even for small scale images, provided that features are kept to a small size value of the order of the width of the PSF of the system. When testing larger features such as a square test pattern we have observed an irreducible speckle-like noise which makes lithographic exposure more difficult. We have seen in these examples how the use of an analytical seed hologram to initiate the process can be used when considering thin patterns, and how this can slightly improve contrast quality in some cases.

We have defined how we can extend the range of depth of our iterative propagations by using a windowing process similar to that of Sypek [20] and also conducted an experiment to demonstrate how this windowing process causes a disturbance to the iterative scheme. Nonetheless, final image quality has been relatively unaffected and we find this extension very useful for working on real scale problems within reasonable simulation sizes.

We have then shown an example of how subsampling of an iterated pattern can produce false positive results and have experimented with two methods (Filtering and Pixelation) to limit the bandwidth of the simulation so that results remain accurate and the best quality images are produced by the iterative scheme.

We have shown an example of how size limiting of a hologram can have a negative affect on image quality and discussed a mathematical limit on minimum hologram size as compared to image size which can give us the best image quality.

Examining an experiment which varies intended feature width we have shown how intended thin patterns smaller than the width of the PSF of the system are

simply expanded to the PSF. Further to this we show how patterns over a certain proportion of the width of the PSF become destroyed by speckle-like noise.

Looking at and experimenting with the background value set on our intended amplitude patterns, we have observed that small background values can significantly improve the quality of the image with regards to contrast and allow the iterative optimisation to not develop significant noise patterns, though yet larger values merely increase background noise.

We have simulated a simple method for the removal or improvement of noise by signal averaging over time during an exposure on an active modulation device and very briefly discussed alternatives such as reduced coherence.

Lastly - taking into account all of our previous limitations and modifications - we have shown an experimental verification of an iterated hologram implemented in an SLM exposure system and compared it to our simulated intensity pattern. As expected we observed an increase in background noise level for the experimental intensity, however the pattern remains well-defined and suitable for lithography.

Chapter 6

An Iterative Algorithm for Lithography in 3D

6.1 Introduction

From the investigation described in the previous chapter, we have derived a suitable approach for iteratively designing holograms, taking into account an application in photolithography by considering a contrast metric and exploring various image and set-up parameters. We have performed experiments to demonstrate how parameters of the system and simulation affect the image output. We have yet to apply these methods to the creation of a suitable 3D image.

Our previous chapters describe a planar iterative algorithm with a single input and output plane. This approach, when using a convolution based transform and transfer function, readily extends to design 3D light constructions. This chapter first describes a naive iteration algorithm, such as that of Makowski [38] which allows for 3D light pattern formation. We then move on to look at the critical issues with this method; namely the application of neighbouring constraints and a partial constraint solution to this problem. Furthermore, we examine these modifications to ensure that the algorithm remains stable and convergent.

All images and demonstrations in this chapter are simulations except in section 6.7 where methods are experimentally verified.

6.2 Multi-Plane Iteration

As before, an iteration algorithm can be formed by considering any one of the propagation transforms described in Chapter 2.

The extended approach proposed by Makowski et al [38] and Dorche et al [37]

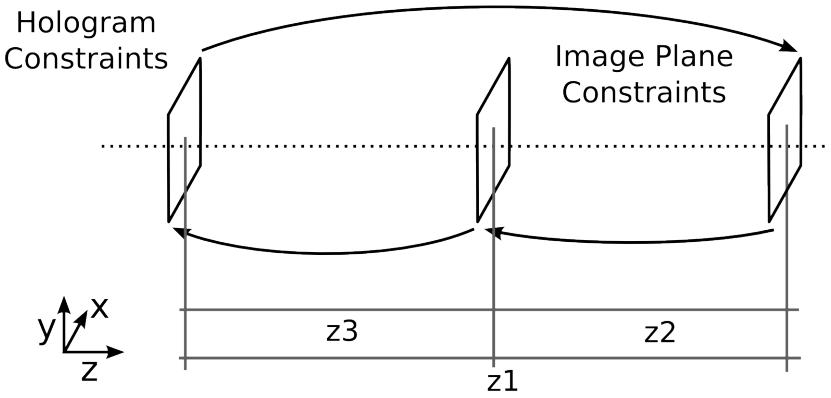


Figure 6.1: Graphical representation of known basic iterative optimisation scheme

to a multi-plane optimisation problem can be used to produce a hologram under constraints which can then be used to generate a 3D image field. The process (similar to the Gerchberg-Saxton algorithm) is comprised of

1. Initialise a random phase field. $U(x,y) = exp(2\pi i * \text{Rand}(x,y)$. Where $U(x,y)$ represents out sampled field in the hologram plane.

2. Multiply field angles by a hologram amplitude constraint pattern (Top hat function for phase-only). $U'(x,y) = \text{Angle}(U(x,y)) \times C_H(x,y)$, where C_H is a sampled top hat amplitude function

3. Propagate forward to an image plane by convolution with transfer function (h) for a given z distance. $I_1(x,y) = U'(x,y) * h(x,y,z_1)$

4. Apply an image amplitude constraint pattern. $I'_1(x,y) = \text{Angle}(I_1(x,y)) \times C_{I1}(x,y)$ Where C_{I1} is the constraint pattern to be applied on the 1^{st} constraint plane.

5. Propagate a small distance to another image plane $I_2(x,y) = I'_1(x,y) * h(x,y,-z_2)$

6. Apply a different amplitude constraint pattern. $I'_2(x,y) = \text{Angle}(I_2(x,y)) \times C_{I2}(x,y)$

7. Propagate back to the hologram plane. $U(x,y) = I'_2(x,y) * h(x,y,-z_3)$

8. Repeat process from step (2).

This algorithm is an extension of the planar algorithm described in section 5.3. The algorithm is outlined graphically in figures 6.1 and 6.2.

As we have demonstrated, certain aspects of this iterative scheme can already be improved to ensure the validity of the optical field. In summary:

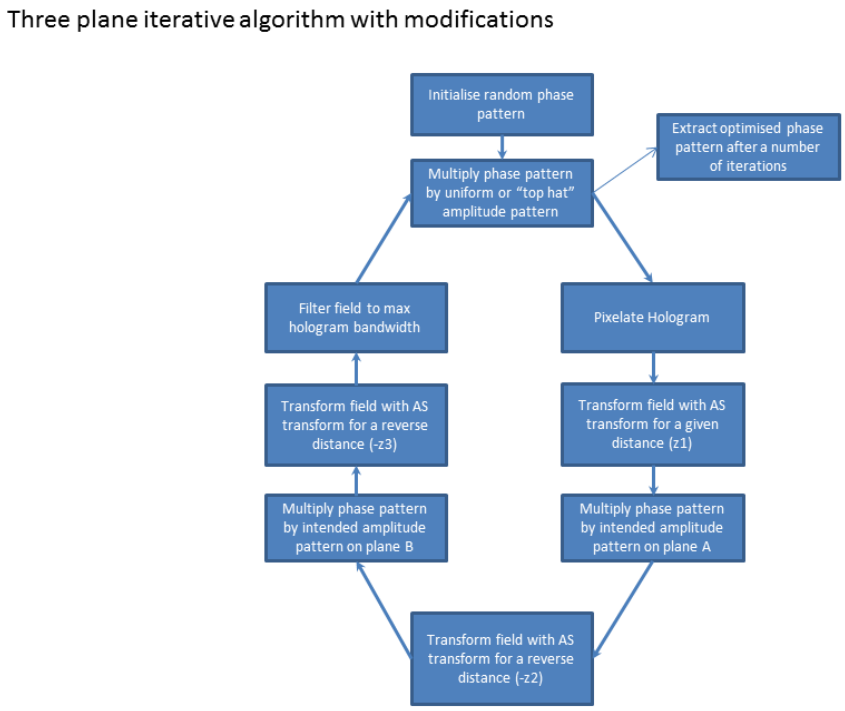


Figure 6.2: Flow diagram outlining the key steps in a three-plane iterative algorithm

1. By adding a filter to the iterative scheme before the hologram amplitude constraint we ensure that the final pattern does not have a greater bandwidth than the modulator or the Nyquist rate of the desired intensity pattern.

2. By applying our pixelation re-sampling function we gain the ability to improve image quality further by more closely approximating how the function is implemented in a modern active modulator.

3. Larger low-noise exposures can be achieved by repeating the iterative process with different random seeds, or by use of a ground glass diffuser, or by improving the coherence of an existing exposure system.

4. We can potentially create useful patterns without the need for partial coherence or signal averaging by limiting ourselves to thin features and carefully considering seed functions.

5. By ensuring that the size of the hologram is sufficient we can maximise the opportunity for the optimisation algorithm to produce a high contrast image, and reduce the development of noise.

We will test and show that using seeding patterns in conjunction with careful control of background value can result in thin continuous features in 3D as well as in 2D.

6.3 Multiple Planar Constraints

Initially it may seem trivial to extend our 2D methods to arbitrary complexity in 3D by considering more and more constraint planes separated along the optical axis.

However, when we try to bring two constraint planes into close proximity some "cross-talk" is observed, and eventually the image becomes completely destroyed by this effect. The work of [38] noted this limitation but the analysis in this paper was insufficient to draw conclusions for our application in lithography, therefore here we repeat a similar set of disparate constraints to demonstrate the limitations of this method and how it may be applied using our well-analysed planar iterative approach.

6.3.1 Full Plane Constraint Limitations

To demonstrate and overcome these constraint issues we start by considering two constraint planes with binary image constraints as shown in Figure 6.3.

We set up a simulation of two binary pattern constraints which are formed from the crest patterns of Durham and Sheffield Universities.

Propagation distance (z)	Wavelength (λ)	Sample/Pixel Pitch
$16/18/32cm$	$405nm$	$4\mu m$
Image Width/Height	**Simulation Width/Height**	**Feature Width**
2048 pixels	4096 pixels	Approx $8\mu m$
Constraint Low Value	**Constraint High Value**	**Number of Iterations**
0.1	1.0	100

Table 6.1: Simulation parameters for multi-plane constraint tests.

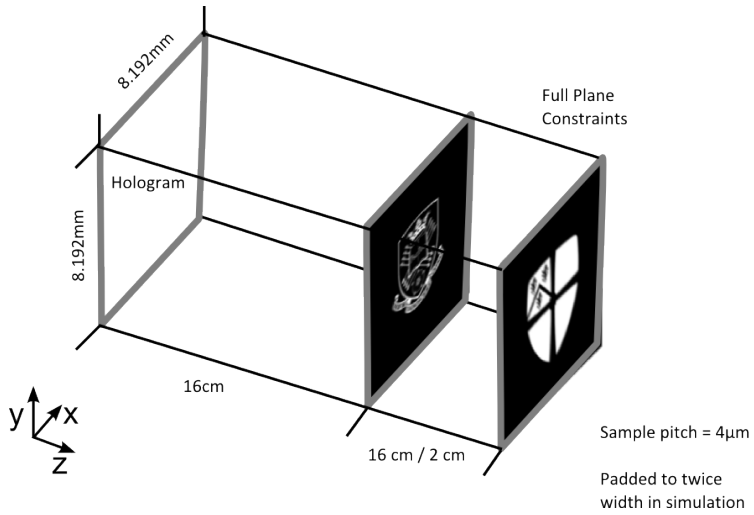


Figure 6.3: Two plane crosstalk simulation setup.

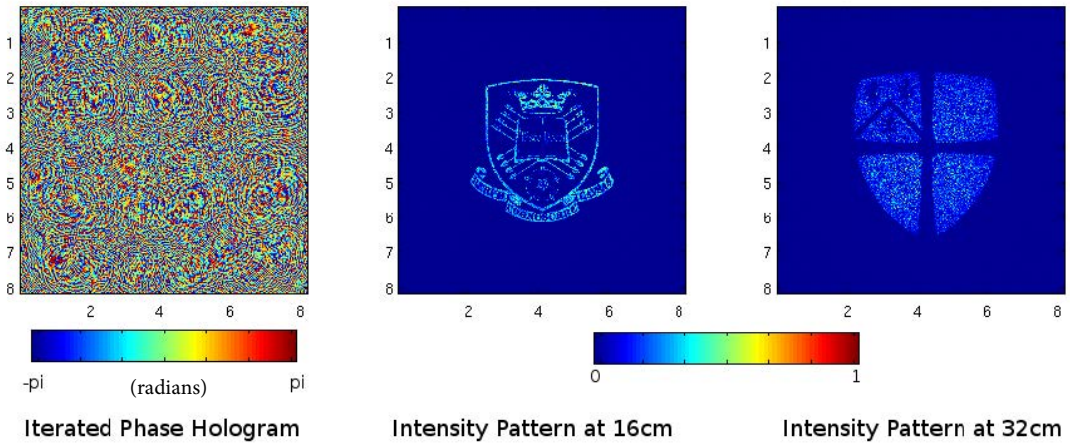
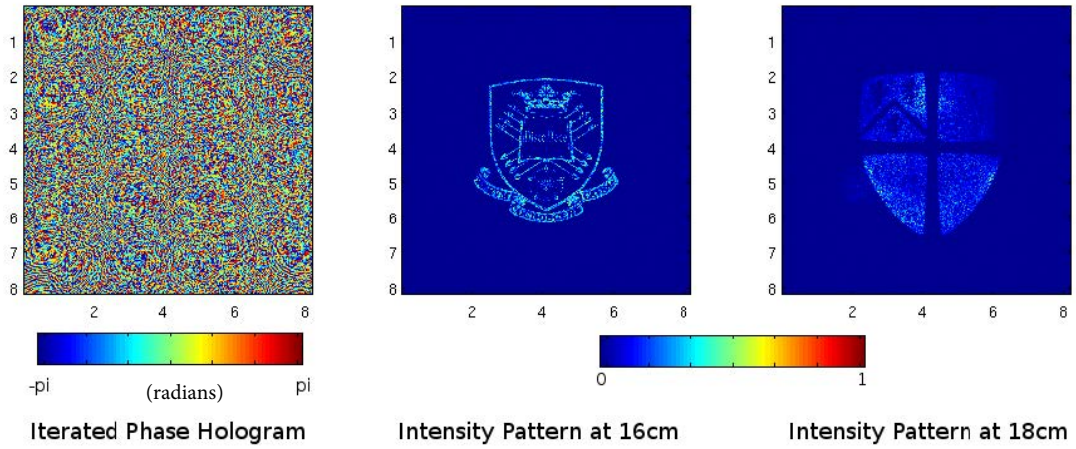


Figure 6.4: Optimised hologram with two simulated intensity patterns at z=16cm and 32cm.



(Spatial dimensions in mm)

Figure 6.5: Optimised hologram with two simulated intensity patterns at z=16cm and 18cm.

Taking the multi-plane method described and applying our own modifications of filtered AS proportion kernel and sampling constraints, we set up two image planes with binary images set at values 1.0 and 0.1 amplitude (schematic shown in Figure 6.3). A random initial phase seed is used.

At 16cm plane separation (Figure 6.4) we find that with sufficient separation, both patterns are well-defined and visible and observe our familiar noise in 'high' regions.

As the image plane constraints are brought closer together (2cm plane separation, Figure 6.5) the solution found by our iterative scheme worsens. Areas of constraint on one image shadow areas on the other pattern causing large variations in intensity across the second image pattern in this case.

This can be understood by considering the validity of the field being created. When two fully constrained field amplitude patterns are placed as constraints close together in space there is no guarantee that a solution exists under the terms of the scalar wave equation, or equivalently the band-limited spectrum of the field (Ewald sphere section). Therefore it becomes much more difficult to find a phase-only solution which closely matches the required image intensity distribution. As constraints are separated further and further along the optical axis, energy has more and more opportunity to spread in the transverse plane. This makes it easier for the iterative algorithm to find an appropriate phase solution.

With simpler and more obviously disparate constraints as shown in Figures 6.6 and 6.7 of a wide bar in either vertical or horizontal orientation we see that both low region constraints in near and far planes have a negative affect on the neighbouring constraint image pattern. This is evidenced by the shortening of the bar in the 2cm case. Only the central region in which both contain 'high' constraint regions is well-defined due to the localised in energy in the transverse plane.

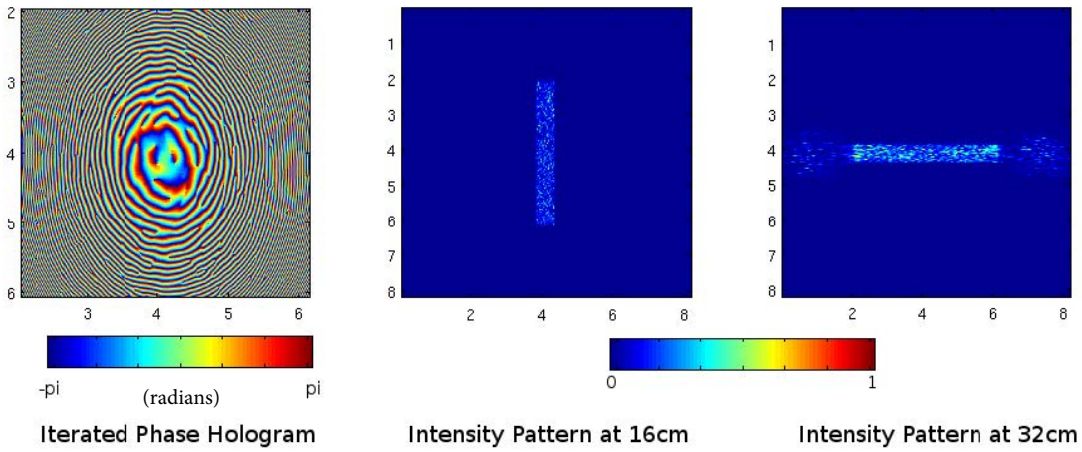


Figure 6.6: Optimised hologram with two simulated intensity patterns of "bar feature" well separated in z at 16cm and 32cm.

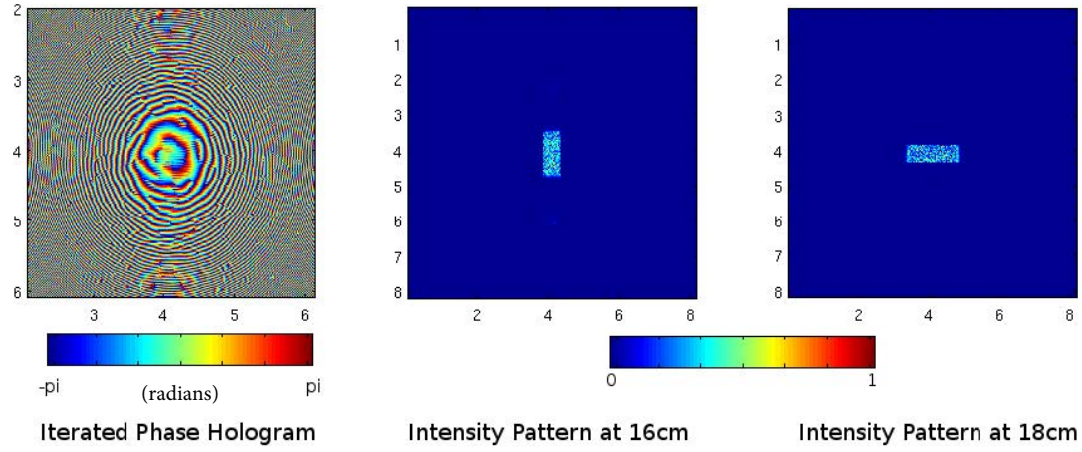


Figure 6.7: Optimised hologram with two simulated intensity patterns of "bar feature" poorly separated in z at 16cm and 18cm. Note the severely reduced length compared to the well-separated constraints

This problem can also be noted in [43] wherein a full constraint is applied to a field simultaneously, resulting in only a rough outline of the object, and much out of plane energy.

6.3.2 Partial Constraints

The work of Xia [40] introduces the concept of masking parts of the image for constraint in some planes and not in others. This process takes each planar constraint and divides it into two areas. One area is constrained to a new image pattern. The other remains unchanged.

By applying the analysis in the previous chapter, the example given in Xia is physically inaccurate due to sampling constraints for the pattern. When simulated correctly with realistic constraints (using the methods described in chapter 5), we observe our familiar noise patterns for phase-only constrained "GS optimised" holograms with a random seed pattern.

Taking the idea of partial constraints further, we consider breaking an image volume into enough planes to adequately represent a 3D surface that we want to pattern. An example for a sloped surface is shown in Figure 6.8. The core idea here is to generate a set of discretised transverse constraints along the optical axis. These partial constraints only apply their constraint pattern to the regions which are close to the surface we wish to pattern i.e. for a sloped pattern in one dimension we produce a set of stepped constraints. This way we construct a quantised volume image with a given sample rate in all directions. An overview of a modified algorithm for applying such partial constraints can be seen in figure 6.9.

This method requires a plan view image of the pattern to be broken into mutually exclusive constraint regions which form a quantised surface. The surfaces which can be generated under this approach are limited by sampling constraints. The surface must also satisfy the constraints of a discontinuous function, which is to say that it cannot double back upon itself or take multiple values at a single x-y coordinate. This is largely irrelevant when considering the kind of surfaces which we would like to pattern. These more extreme surfaces would introduce the problem of complete occlusion to parts of the substrate. This constraint will be discussed in section 6.6.1.

By adopting this method, we will show that it is possible to attenuate the issue of 'crosstalk' between image plane constraints. This is because information is not occluded or altered in propagation through unconstrained parts of transverse constraints.

Several questions about this approach arise: By the adoption of a partial constraint technique, do we break any serious assumptions made about the iterative

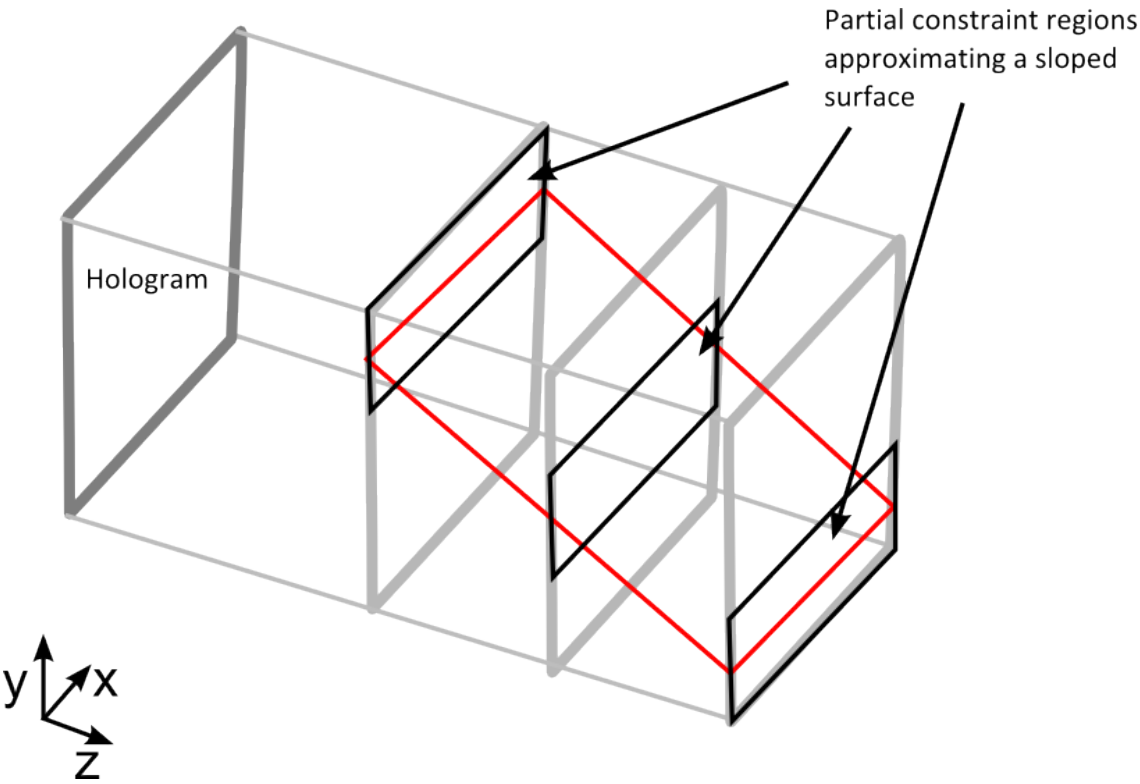


Figure 6.8: Example constraint discretisation

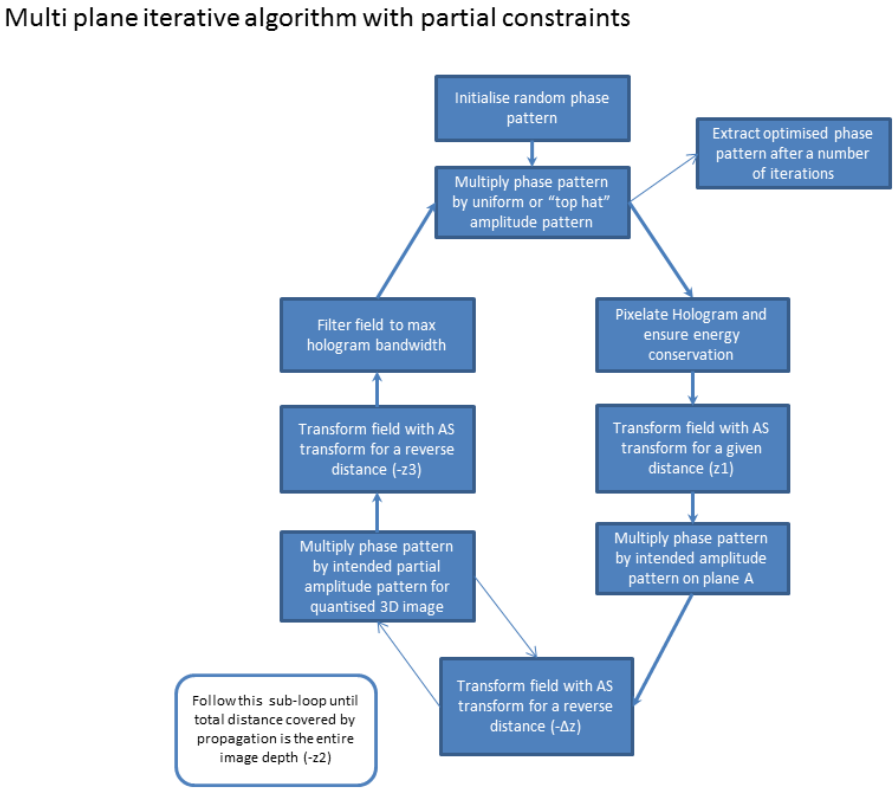


Figure 6.9: Overview of multi-plane partial constraint algorithm.

algorithm and its error reduction? What level of constraint is required to form a continuous pattern on a 3D surface? Finally, is the increase in computational burden viable?

6.3.3 Multi-plane convergence issues

When applying partial plane constraints it would be easy to ignore the effect of the partial constraint, and assume that the system behaves much the same as a single plane constraint, or a set of complete plane constraints. This, however, is not the case, and furthermore has serious consequences to the final image pattern.

Energy Conservation and the GS and Multi-Plane Approaches

As stated in the original paper by Gerchberg and Saxton, their iterative algorithm relies on Parseval's theorem to ensure that the integral of the field either side of the FT is equal and therefore that the error in the phase must always decrease when applying amplitude constraints to both planes. Energy conservation is maintained when moving to a free space convolution propagation (FDF, FC, AS), since the transfer function is a phase-only function. On each constraint plane a full-field amplitude constraint is applied making it impossible for the total amount of energy in the system to change between these constraints.

With a multi plane partial constraint pattern, however, this energy conservation can no longer be assumed.

One issue with this assumption of energy conservation of the system is that, problematically, the basis of the AS iterative algorithm is non-physical. The angular spectrum transformation energy which diffracts outside of the image field is wrapped back into the field on the other side due to the cyclic nature of convolutions performed with FFTs. Windowing, as discussed in section 5.8, alleviates this problem but causes other slight errors which we have examined. Assuming that most energy is confined to the image field and not removed by the windowing process (a reasonable assumption as we are attempting to form a complete pattern on the inside of the image constraint region), we have in effect a closed system, as our error reduction algorithm requires.

The iterative convergence begins to break down when we consider 'large' (disproportionately to the sum of energy in the image constraints) amounts of unconstrained energy passing through our iterative constraints. Since so far all of our constraints have been field amplitude patterns applied to the phase pattern in our numerical simulations, the opportunity exists for energy to remain unconstrained and unattenuated through the entire depth of the image.

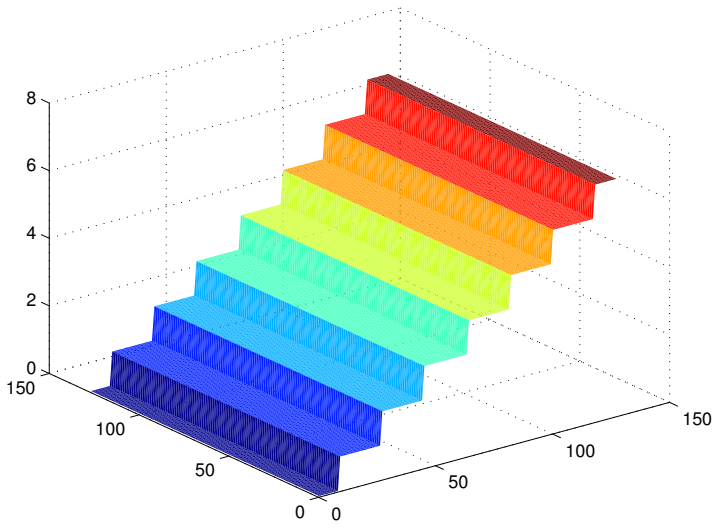


Figure 6.10: A depth quantised slope pattern showing boundaries for each depth region.

Specifically in the case of phase-only holograms this can become a more serious issue because of the application of the phase-only constraint at the hologram plane. The phase-only constraint is implemented by setting the amplitude of the field to a constant over the entire hologram plane, i.e. no low region. This can cause an issue when propagated back to the image constraints, as it potentially constitutes a large gain in energy when compared to the image pattern itself, which then has the potential to propagate back through the entire system unconstrained. The consequence of this lack of constraint is a failure of the iterative algorithm to converge upon the desired image pattern and instead to show large intensity patterns in unconstrained areas.

Figure 6.11 shows a subsection of a horizontal line patten which should focus on a quantised slope (shown in figure 6.10). This pattern when broken into few separate constraint planes develops intense regions at the constraint boundaries (i.e. the vertical lines shown in the image), at the expense of the intended pattern. Parameters for these simulations are given in table 6.2.

A simple diffraction efficiency metric is used to analyse the pattern for comparison. We find that 1.72% of the energy in the field is diffracted into the pattern region.

This issue could be circumvented in several ways. One possibility is to modify the iterative algorithm such that information in regions occluded by our quantised surface is discarded before each propagation. However, this makes a single pass of

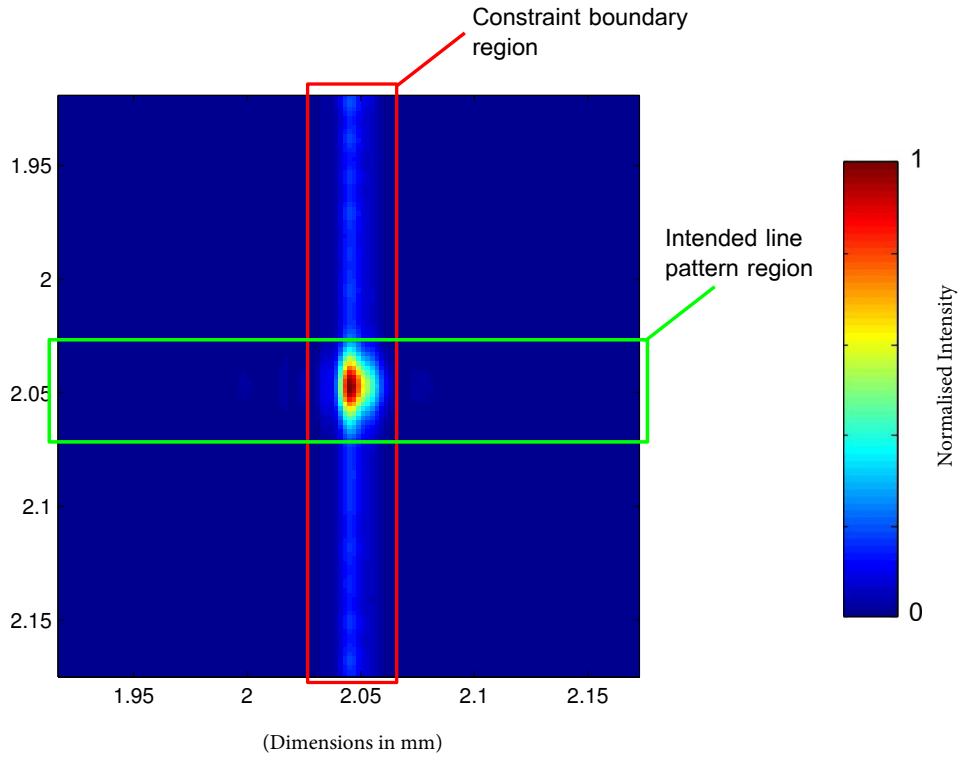


Figure 6.11: Iteration with unsuppressed unconstrained noise after 50 iterations. Showing intended line pattern region.

Propagation distance (z)	Wavelength (λ)	Sample/Pixel Pitch
$8cm$	$405nm$	$2\mu m$
Image Width/Height	**Simulation Width/Height**	**Feature Width**
2048 pixels	4096 pixels	Approx $4\mu m$
Constraint Low Value	**Constraint High Value**	**Number of Iterations**
0.05	1.0	50

Table 6.2: Simulation parameters for multi-plane partial constraint simulations, with and without energy conservation modifications.

our image field impossible; information with regards to large parts of the field would be discarded from the occluded regions. Therefore, after propagating the field back to the hologram, the field propagation would be required to be reversed and the phase calculated on each constraint section with no constraints applied.

Setting to zero all regions which are occluded by the surface means that - as long as we were to consider a complete surface - no region would be unconstrained under the surface. This would mean no stray energy in the system, and therefore remove our issue of unconstrained energy at constraint boundaries. However, by taking some care over how constraint patterns are applied, this algorithm need not be lengthened.

Instead of applying a unit amplitude pattern reference at the hologram plane and not considering the energy of the system (which can lead to this constraint greatly increasing the energy in the system) we simply ensure that we apply a constant amplitude pattern where the energy before and after the constraint remain constant (i.e. we scale this top hat amplitude function which is applied to maintain a constant sum of squares). This is shown by the simple snippet in listing 6.1. This means that no energy is added to the system in this single full plane constraint and therefore, the image pattern is not swamped by unconstrained energy.

Listing 6.1: A simple phase-only energy equalisation constraint for sampled field P.

```
1 energy = SumOfSquares(P);
2 P=P./abs(P);
3 energy_new = SumOfSquares(P);
4 scale_factor = energy/energy_new;
5 P = P.*sqrt(scale_factor);
```

Figure 6.12 shows the same pattern as shown in figure 6.11, with the energy equalisation modification applied each time the phase-only constraint is applied to the hologram. In this case it is plain to see that the pattern image is significantly improved. Quantifiably, our simple diffraction efficiency metric shows 14.7% of the energy is now within the pattern region. This is an 8.5 times improvement on our previous 1.72% efficiency. You will note that end sections of the line pattern appear to be out of focus. This is because the image shown is that of a focal plane in the centre of the image field.

6.3.4 Image contrast and constraint plane separation

Now we turn to examine how the inter-plane separation of our partial constraint algorithm affects the final image quality. Several questions must be answered with

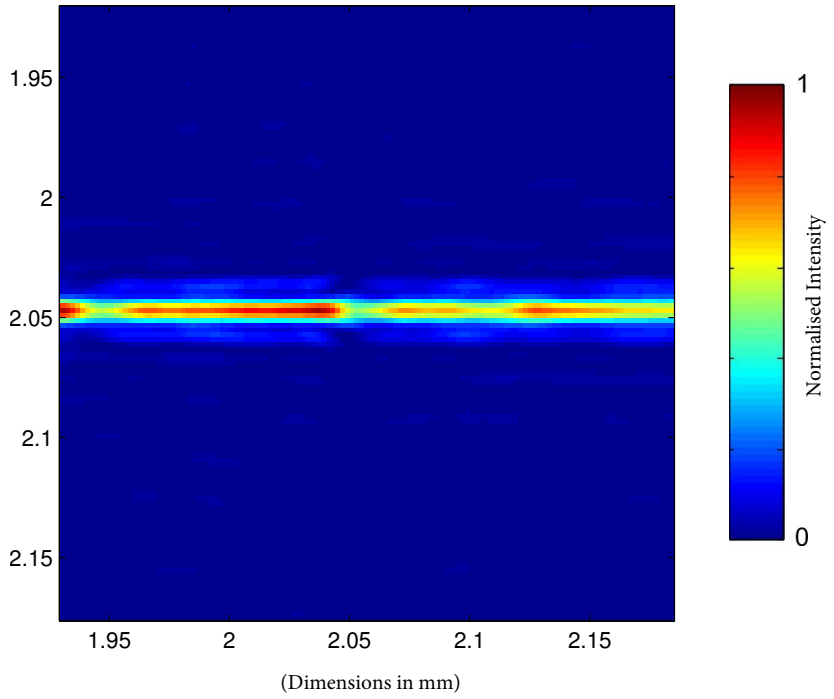


Figure 6.12: Iterated image showing no boundary noise due to energy equalisation suppression. Two line sections constraints are shown here focussed onto an intermediate plane, and so are slightly out of focus.

regards this approach. Firstly, is it possible to use multiple partial constraints to form a continuous pattern? Secondly, at what separation distance do we expect to be able to see "continuous" features with sufficient contrast to be used in lithographic exposure? Thirdly, how does breaking a line into many multiples of partial constraints affect the output image, i.e is the noise generated by having discrete constraint sections prevent the use of the method?

For a low sample rate in 'z', i.e. few/sparse constraint planes, we expect significant discontinuities in the image. In this case it is accurate to consider this system as several separate planar images. Figure 6.13 shows an example of a line pattern in focus on one constraint plane, and the beginning of another out of focus section on another constraint plane.

As the number of planes increases and/or the gap between image planes decreases contiguous exposure regions can potentially begin to form.

Three experiments are presented to address this multi plane constraint algorithm.

Firstly, we show a simulation of two zone plate holograms where the focal points are separated by varying distances in x and z. This allows us to gauge an analytical approximation for constraint depth separation versus contrast.

Secondly, we generate a split line pattern on two constraint planes which allows

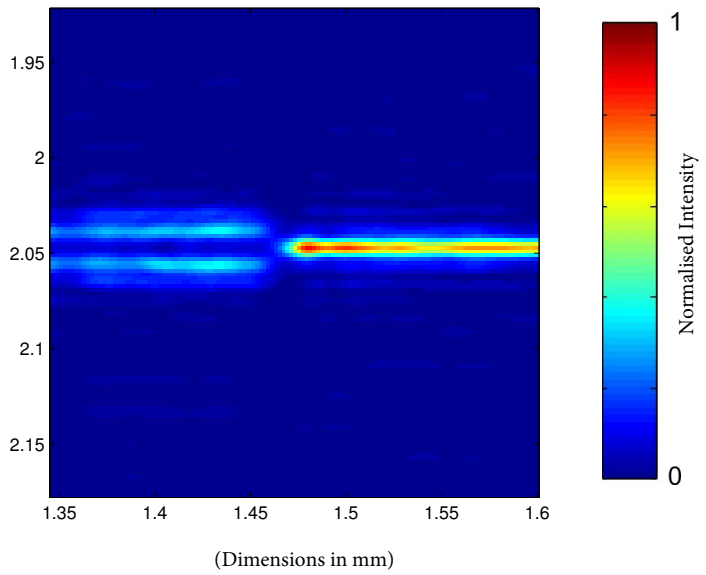


Figure 6.13: Iterated image showing line image split across two constraint planes at different depths.

Propagation distance (z)	Wavelength (λ)	Sample/Pixel Pitch
$8cm$	$405nm$	$2\mu m$
Image Width/Height	**Simulation Width/Height**	**Feature Width**
2048 pixels	4096 pixels	Approx $4\mu m$
Constraint Low Value	**Constraint High Value**	**Number of Iterations**
0.1	1.0	50

Table 6.3: Simulation parameters for the following multi-point, dual-plane and multi-plane experiments

us to examine how varying inter-plane separation affects the resultant continuity of the image for a simple 2-plane constraint.

Finally, we set up a 'continuous' line on a tilted plane. When varying the inter-plane separation we look for a minimum viable plane separation in the particular example of a 45 degree tilted plane. Simulation parameters used throughout these next experiments are summarised in table 6.3

Inter-plane separation experiments related to depth of focus

As noted previously the depth of focus of the optical system is controlled by the numerical aperture of the focussing element or limiting aperture, or equivalently the bandwidth of the optical system, which as we have investigated may well be dependent upon the pixel density of the optical modulator. The depth of focus of

the optical system is therefore governed by the equation:

$$\delta z = \frac{\lambda}{1 - \cos(\theta)} \tag{6.1}$$

where

$$\theta = \arcsin(\text{NA}) \text{ in air} \tag{6.2}$$

Note that for a filtered hologram we derive an effective NA from the maximum bandwidth of the hologram:

$$\text{NA}_{\text{effective}} = \nu_{\max}/\lambda \tag{6.3}$$

$$\nu_{\max} = \frac{p}{\delta x} \tag{6.4}$$

where p is the proportional size (compared to the maximum simulation bandwidth) of a circular filter applied to the spectrum of the field and $\frac{1}{\delta x}$ gives $\nu_{\max}$ for a sampled system at δx sample pitch.

To form a continuous pattern in x-y a classical approach would be to define that two points must be indistinguishable with regards the Rayleigh criterion which is related to the bandwidth of the pattern in x-y. We have taken this slightly further in our planar simulations. It is therefore sensible to assume that to achieve effective "continuity" in z we must separate adjacent 'pixels' by some proportion on the order of the depth of the point spread function as determined by the optical bandwidth.

Analytical point hologram separation experiment

Taking a phase-only zone plate point hologram, we simulate a slice of the field generated along the y-z axis. This pattern is then added to a duplicate of this field slice, where each sample has been phase shifted by a fixed phase factor. Schematically, this is shown in figure 6.14. This process of superposition of two zone plate point spread functions models the superposition of two phase and amplitude zone plate holograms. By moving the two PSF foci relative to each other and adjusting the phase of each function we are able to gauge how 'distinguished' each point is from the other for various separations.

The optimal phase factor is determined by calculating the intensity at the average position of the two focal points. Maximising the intensity at this centroid by a brute force test of varying phase gives us the best possible contrast when attempting to blend the two individual points into one continuous point.

By varying the position of one of the focal points in both x and z we are able to determine at what separation we are no longer able to form a continuous 'line'

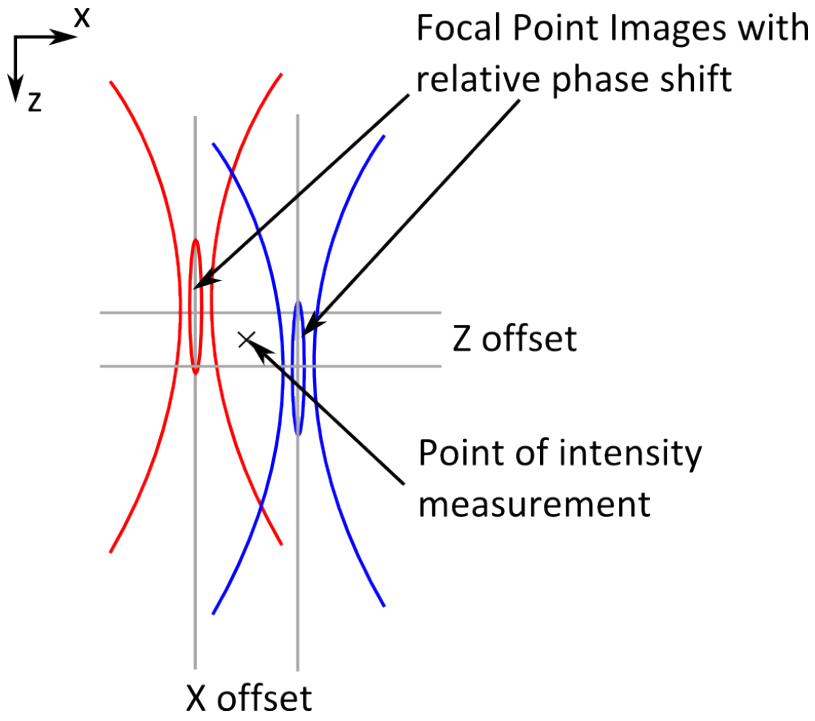


Figure 6.14: Two point continuity simulation experiment setup. A sample pitch of $2\mu m$ is used in the x direction and $20\mu m$ in the z direction. x-y feature size is limited to approximately $8\mu m$ by a circular filter. DOF of the PSF of the system is approximately $1.3mm$

from the two separate focal points. The intensity of this centroid point for a range of values of Z offset at an example fixed X offset is shown in figure 6.15.

For later comparison with iterative approaches we add a range of background intensity values ranging from 0.005 to 0.5 as a proportion of the maximum intensity to put this intensity curve through our contrast measure as the 'high' value with background intensity as the 'low' value. The results are shown in figure 6.16.

We now have an approximation showing that we expect to see contrast in our line pattern deteriorate for plane separations between approximately 1 and $2mm$ plane separation, where the effective 'point' spacing is $2\mu m$ for a bandwidth limited feature size of approximately $8\mu m$. This is not unexpected as the DOF of the optical system is approximately $1.3mm$.

In the next two experiments we will compare these findings to what we achieve with iterative holograms under similar constraints, i.e. band limited holograms with a feature size of $8\mu m$ sampled at a pitch of $2\mu m$.

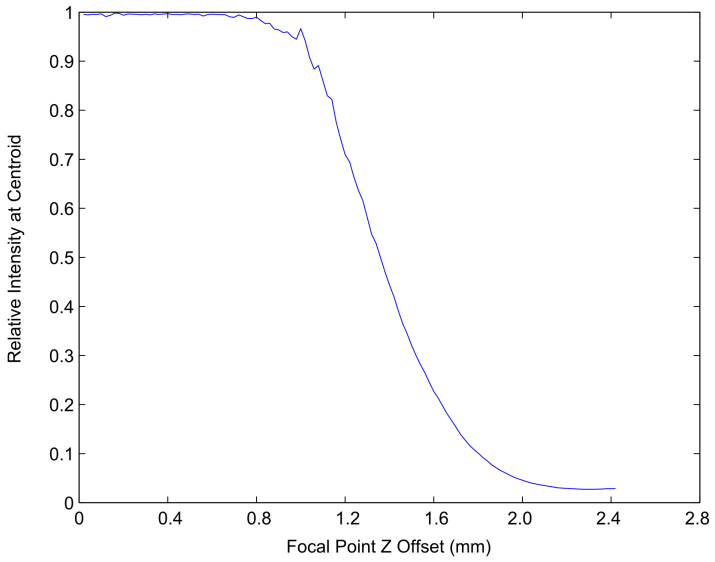


Figure 6.15: Relative centroid intensity as two points separated by $2\mu m$ in x are variably separated in z.

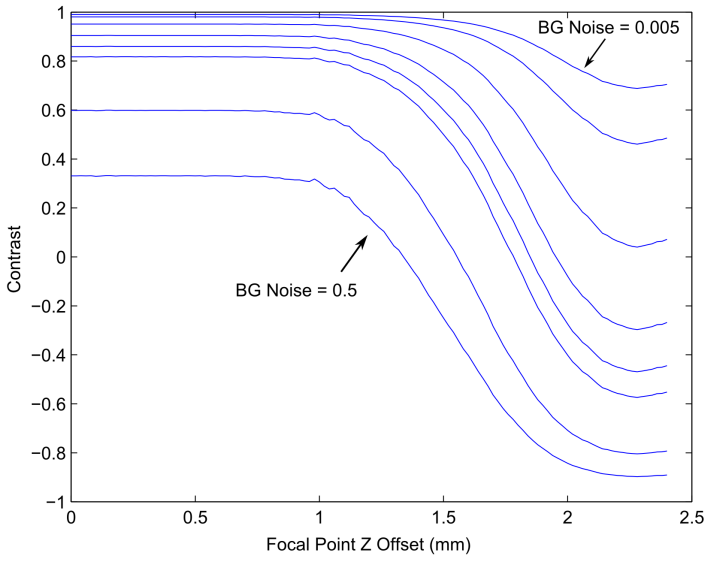


Figure 6.16: Relative centroid contrast as two points separated by $2\mu m$ in x are separated variably in z. Shown is calculated contrast for a variety of values of relative background intensity from 0.005 to 0.5.

6.4 Iterated line segment constraints

Now we look at the simplest possible partially constrained multi-plane iterative case, wherein a line pattern is split across two partial constraints, which are separated by a variable distance (Figure 6.17).

Nominally the image volume containing the two constraint planes is placed at $8cm$ from the modulator. The distance between these two constraint planes is then varied and the final contrast after 50 iterations for a profile along the line image is assessed with our stringent contrast metric, on an intermediate plane which represents where a sloped surface between the two constraint planes would be most out of focus.

To keep our contrast measure accurate, the simulation sample spacing is kept at at least 4 times the maximum frequency supported by the filtered transfer function, which likewise should be a sufficient sample rate for the intensity pattern generated by the field keeping scalloping error to a minimum (despite the increased frequency content introduced by squaring the field).

The effective numerical aperture (and thus DOF) of the system is also varied by changing the filter which is applied to the propagation transform kernel.

It should be noted that since we consider a surface of neighbouring points in x-y which are then also offset in z that this is not exactly the same case as simple spacing in x-y, i.e. points are not directly on top of each other in the z direction. This, however, is the most straightforward approach to partition the image onto a surface by using non-overlapping image constraints.

Such overlapping constraints could make a good extension to this work at a later date. It is nonetheless interesting and practical to see how far we can push planes apart in this regime and still achieve a reasonable continuous pattern.

Figure 6.18 shows the raw data of contrast for a variety of DOF values, from $1.3mm$ to $5.1mm$ achieved by filtering the hologram. The inter-plane separation of the constraints is varied from $0.0mm$ to $5.0mm$ in increments of $20\mu m$. The contrast of the line pattern is obtained by analysing the focal image at a plane halfway between both constraint planes. This represents the 'worst case' focal image for a sloped line quantised into two planar constraints, as the central section of the slope will be exposed at this focal plane. A profile is taken along the line image on this intermediate plane and divided into 'high' and 'low' regions with a 50 pixel gap at the end of the line to account for possible edge effects negatively affecting the measure.

The raw data shows several trends, primarily the decrease in contrast as the interplane separation increases as expected. Moreover the increase in sensitivity to

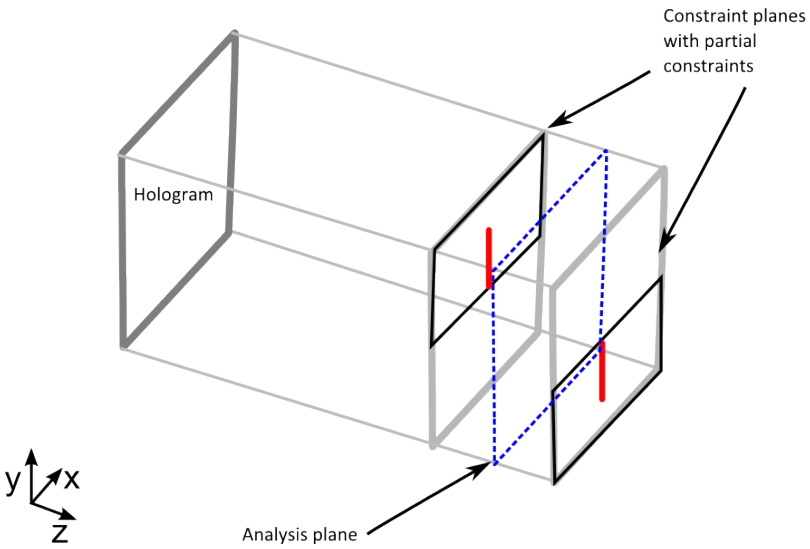


Figure 6.17: Line image and constraint patterns

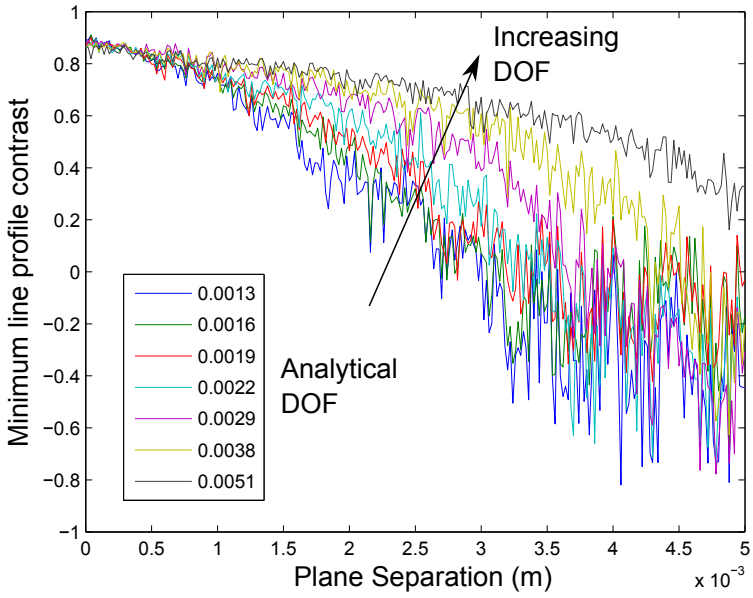


Figure 6.18: Single line pattern contrast after 50 iterations for varying effective DOF, against partial constraint plane separation for a two plane line image on a simple slope.

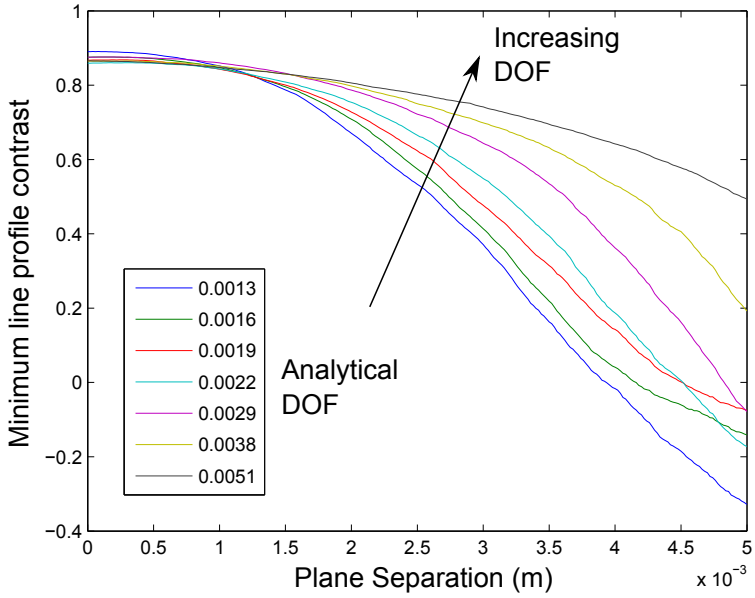


Figure 6.19: Contrast data as shown in Figure 6.18 conditioned with a simple moving average filter. Clearly demonstrating the underlying trend.

plane separation as the DOF decreases, again as expected. These trends are shown more clearly in a moving-average-filtered version of the graph shown in figure 6.19.

It should be noted that an increase in variation is also observed as we increase the plane separation, this is potentially linked to the fact that as the contrast decreases this measure becomes more sensitive to background noise level, though other factors likely also contribute.

Comparing this graph to the previous two point analytical continuity contrast graph (figure 6.16), 0.0013m (the blue line) in figure 6.19 matches the simulation conditions of the two point contrast experiment. We find that the roll off of this function is more gradual than that shown by our initial approximation. In the two sectional plane constraint case (for DOF of $1.3mm$), contrast reaches 0 at approximately 3 to $3.5mm$ (or approx 2.3 to 2.7 times DOF). In the two point case (Figure 6.16) this critical point (for an equivalent maximum contrast of approx 0.8) this value is between 1.5 and $2mm$ (1.2 to 1.5 times the DOF).

This more gradual reduction in contrast is potentially understood by considering symmetrical reinforcement in the image. It can be seen with an analytical line hologram that symmetry across the line axis reinforces the pattern, along its x direction. Provided that the phase pattern is correctly adjusted, each section of line hologram forms a sinc function which interferes constructively with neighbouring segments. We suggest that a similar effect is occurring in this iterative case allowing increase of the contrast parameter to form a continuous line where in the simplest

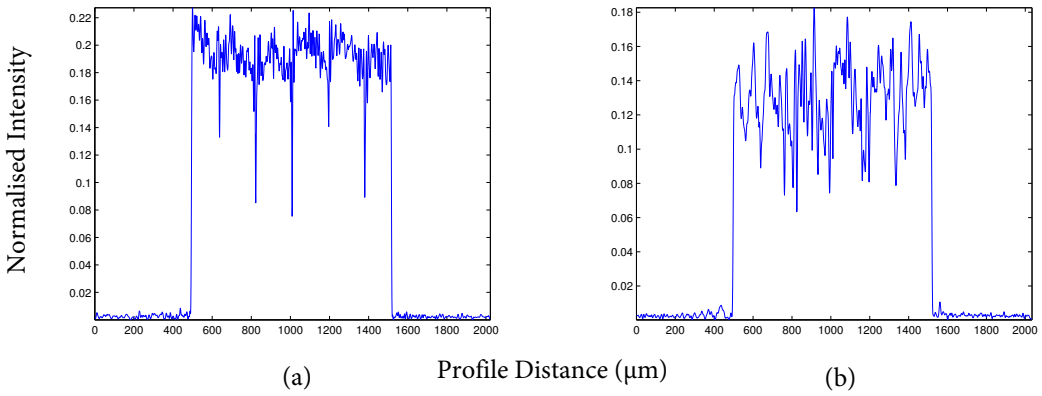


Figure 6.20: 45 Degree sloping surface pattern for $\delta z = 409.6\mu m$ (a) planar phase seed , (b) analytical phase seed. Accumulated from focal images on 2048 focal planes on the sloped surface each one pixel wide.

two-point case we would expect a discontinuity.

In conclusion from this experiment we observe that the inter-plane separation need not necessarily be as strictly constrained as observed in our initial two point experiment. As we will go on to see, however, this contrast parameter is highly pattern dependent.

Sloped single line example

In our third test of 3D line and point patterns, we attempt to create a full-sloped line focal image by breaking our image volume into many quantised constraint planes, and varying the inter-plane separation accordingly, allowing us to create a 45 degree sloped line patten.

We test this line pattern for varying inter-plane separation and analyse the resulting image by simulating the image on the sloped target surface. This analysis is done by taking one vertical line of pixels from a set of focal planes intersecting the surface in the image volume. These lines are composited together to create a sloped plane image which can be analysed for line profiles and cross-sections with our contrast metric.

In addition, we have tested both analytical line hologram and plane phase seed patterns as starting conditions for the iterative algorithm.

Figure 6.20 shows sloped surface profiles of our line focal pattern with the inter-plane separation set at $409.6\mu m$ which is 0.32 times the DOF of the optical system ($1264.0\mu m$). The drops in contrast at the inter-constraint boundaries are clear in the case of the analytically seeded pattern, but less obvious in the plane phase seeded pattern, due to a larger variation in the high region of the image. This case has sufficient contrast for both analytical and plane phase seeded holograms as minimum

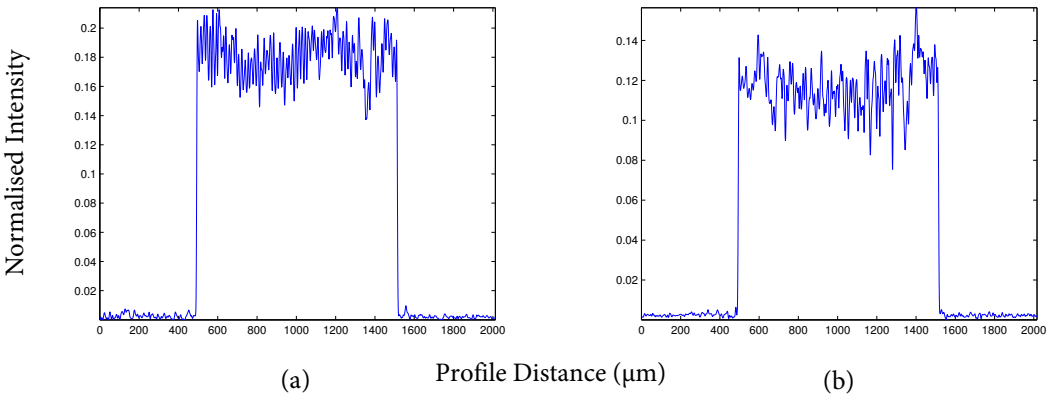


Figure 6.21: 45 Degree sloping surface pattern for $\delta z = 409.6\mu m$ (a) planar phase seed , (b) analytical phase seed . Accumulated from focal images on 2048 focal planes on the sloped surface each one pixel wide.

contrast for both is well above 0.6. However it is interesting to note the larger general variation in the "high" region of the plane phase seeded hologram, whereas with the analytical seed, this variation is minimal making the inter-constraint noise very obvious.

Figure 6.21 shows the same pattern at the much smaller interplane separation of $32.3\mu m$ which is 0.03 times the DOF. Here the drops in contrast at the interplane boundaries have completely disappeared leading to an almost ideal line profile. The plane phase seeded pattern appears to be slightly more variable in the high region but the difference is small enough to be regarded as insignificant.

Figure 6.22 shows a comparison of the expected two point plane separation contrast and the analytical seeded and plane phase seeded contrasts for a range of interplane separations. The two point contrast curve chosen for comparison is a baseline that sees maximum contrast just below the expected practical limit of 0.6.

In addition, comparing the planar phase multi-plane line to the same set-up in our previous two-plane test we note that the required interplane separation has decreased i.e. more dense constraints are required. For comparison we observe that our critical zero-crossing is now at approximately 1× DOF for the seeded case and 1.7× DOF for the plane phase seeded case, both less than the approximately 2.3× DOF we observed in the sectional constraint case. Cyclically, this suggests that increasing the complexity of the constraint pattern, increases the required complexity of the constraint pattern to achieve some contrast threshold.

6.4.1 Interplane Separation and Image Complexity

Noting both that more complex constraint requirements decrease attainable contrast, and also that seed pattern has a heavy effect on the attainable pattern, we

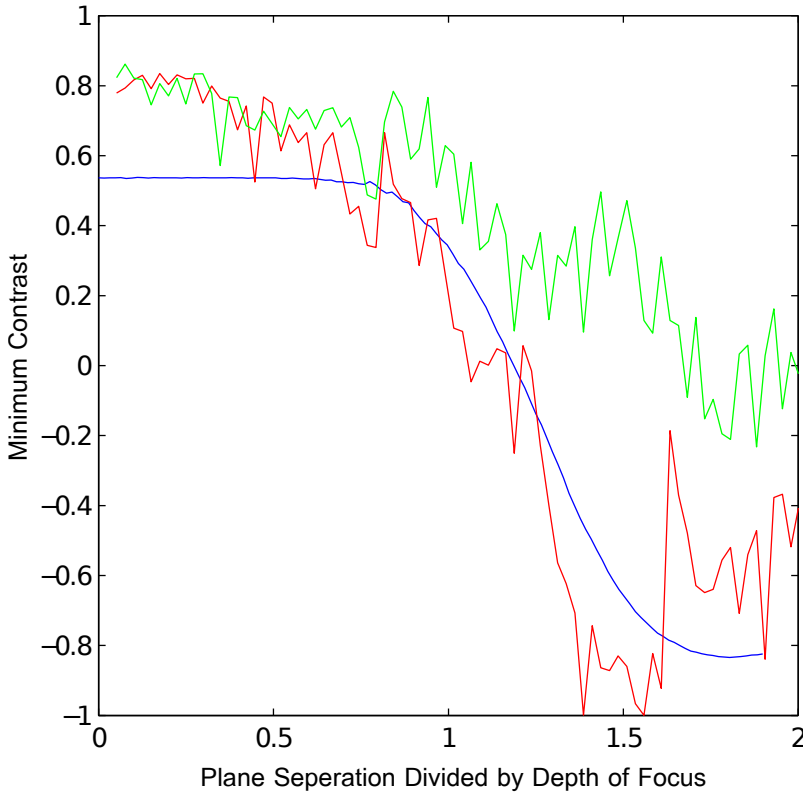


Figure 6.22: Comparison of expected analytical contrast (blue), plane phase seed (green), and analytical sloped line phase seed (red). Contrast for varying interplane separation divided by system depth of focus.

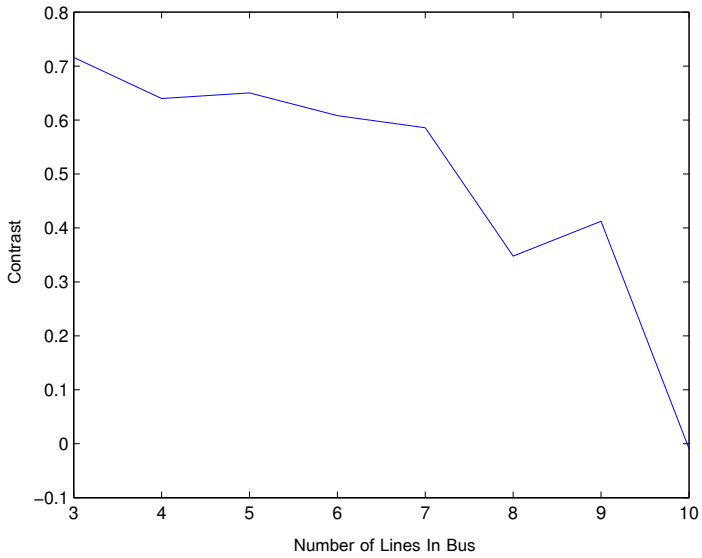


Figure 6.23: Comparison of pattern complexity (number of lines in a dense $32\mu m$ pitch bus) against attained contrast after 50 iterations.

now go on to show that even relatively minor increases in pattern complexity further increase the required constraint burden.

By taking the same set-up as our previous single line experiment, we increase the number of lines whilst maintaining a fixed interplane constraint distance of $758\mu m$ which is on the lower end of the viable interplane separations for a single line. The lines have a pitch of $32\mu m$ which is approximately 8 times the minimum intensity feature size set by the filtered transfer function PSF.

Figure 6.23 shows a graph comparing number of lines used for this setup with the achieved contrast after 50 iterations. Once again we observe a decrease in contrast due to this increased pattern complexity, likely suggesting that an increased number of constraints is required. One final issue to note is that increased pattern complexity is not the only factor in play in this experiment. By increasing the number of lines in the image we are also distributing a fixed amount of input energy over a larger pattern area. This makes it difficult to seperate the effects of complexity against the effects of larger pattern area, but no less valid that these combined effects result in a lower contrast as we have shown for this particular pattern.

6.5 Interplane Separation Conclusions

Though the plane phase seeded iterative image does not match the expected analytical contrast very closely in figure 6.22, similar trends are observed. Both tests

reach a consistent appropriate contrast below 0.6 × DOF. It is interesting and yet unexplained that the plane phase seeded holograms exhibit improved contrast for larger separation where the analytically seeded holograms are not able to achieve the same level of contrast. By comparing these various experiments we have also seen increases in required constraint density from increases in constraint complexity and pattern complexity. For these thin line buses it seems that all of these pattern variations did have seed, pattern and constraint variations that lead to a usefully high contrast.

6.6 Practical issues with 3D patterning

6.6.1 Occlusion

Consider a sloped surface. If the surface is almost flat with regards to the optical axis, we expect that no issue will be caused by occlusion of the wave-front by the surface. For each point on its surface we can draw straight, uninterrupted lines between each point and either edge of a given optical element, be it a hologram or a lens. Now consider a tilted surface which is almost vertical with regards to the optical axis. Here we expect the surface to occlude parts of the wave-front which intend to converge on the focal surface.

A field feature of $8\mu m$ can be formed by a spectrum whose extent reaches $\frac{1}{\delta x}$ or $125000m^{-1}$, which corresponds to a numerical aperture angle of 0.0506 radians for $405nm$ wavelength light. Equivalently this value can be calculated using the grating equation $sin(\theta) = \frac{\lambda}{d}$ or $\theta = sin^{-1}(\frac{\lambda}{d})$ where d is our feature size giving $\theta = 0.0506$ radians again.

Using this angular approximation enables us to define a cone of fixed angle around any given image point where we require that feature size or below. This cone represents where we would expect to see significant contributions from the preceding field. Indeed from our simulated convolution we know that no contributions from higher angled plane waves exist. Figure 6.24 shows an image of three planes of a PSF propagated away from a bandwidth limited point image. It is clear that outside of a central cone angle intensity diminishes significantly. Here we observe that outside of an angle of approximately 3 degrees the pattern diminishes significantly.

This approximation has been used to help us define acceptable limits on the slope of a given lithographic surface. We do this by merely stating that we should not exceed any surface angle that would occlude parts of our bandwidth limited cone around each point.

For $8\mu m$ features, this angle is very small 0.0506 radians or 2.9 degrees, leading

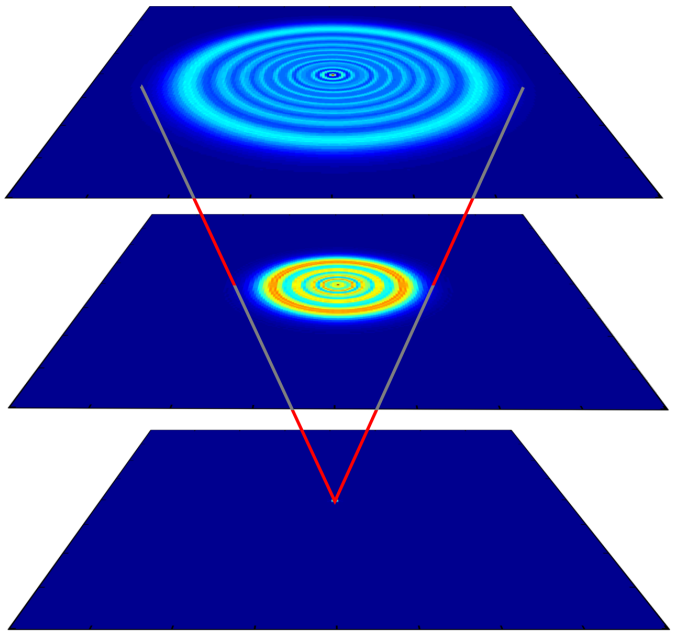


Figure 6.24: The PSF cone intensity formed by a bandwidth filtered point image, on 3 planes.

to a maximum surface angle of 87.1 degrees. All of our examples in this document are far from this limit, aside from those discussed in section 7.4 (Vertical patterning), where a modification is made to allow the breaking of this limit. For this reason this limit has been largely ignored in our considerations. However, it may make an interesting future research topic to attempt to optimise patterns where part of the field is occluded without our further modifications.

6.6.2 Surface angle considerations

Any sloped surface which is illuminated from a given direction experiences a reduction in observed intensity on the surface due to the angle of incidence of the light pattern, as defined by Lambert's cosine law. All of our intensity simulations to this point, assume a flat surface directly normal to the optical axis. It should be noted that this reduction in intensity does not affect our contrast measure for any given section, as noise and signal patterns are affected equally. It does however affect exposure time, as intensity is reduced. Because of this, a surface with varying slope may require that the intended intensity pattern be varied to compensate for the change in energy density. This is possible, and the iteration process will attempt to compensate and brighten some areas with respect to others. However anecdotal experience suggests that this modification might have a severe effect on pattern contrast. Regardless, for the examples produced in this thesis, the minimum 0.6

contrast metric which we have relied upon has proven adequate for exposing any varying surface up to 45 degrees without this compensation, showing little or no impact on exposed feature size. We therefore leave this investigation to further research and development of the method.

6.7 Experimental Verification

Single line and multi-line holograms were implemented on our SLM exposure system and captured on a CMOS sensor with $1.67\mu m$ pixel pitch.

Figure 6.25 shows a multi-plane line with an inter-constraint distance of $758\mu m$ which is approximately 0.6 times the DOF of the optical system, which according to our previous test should be suitably small to keep contrast above 0.6. Two focal images are shown, one with the central constraint plane in focus and one at a focal plane between two constraint planes. It can be seen that it is indeed the case that the intermediate focal image maintains a high intensity at the constraint boundary.

Figure 6.26 shows a bus pattern with the same inter-constraint separation and clearly demonstrates the serious drop in intensity at the interplane constraint boundary. Figure 6.27 shows the same pattern but with the interplane constraint distance dropped to $100\mu m$ which is approximately 1/13th the DOF of the optical system. With this stricter constraint pattern we observe a suitable and high continuous intensity over the many constraint boundaries in the focal regions.

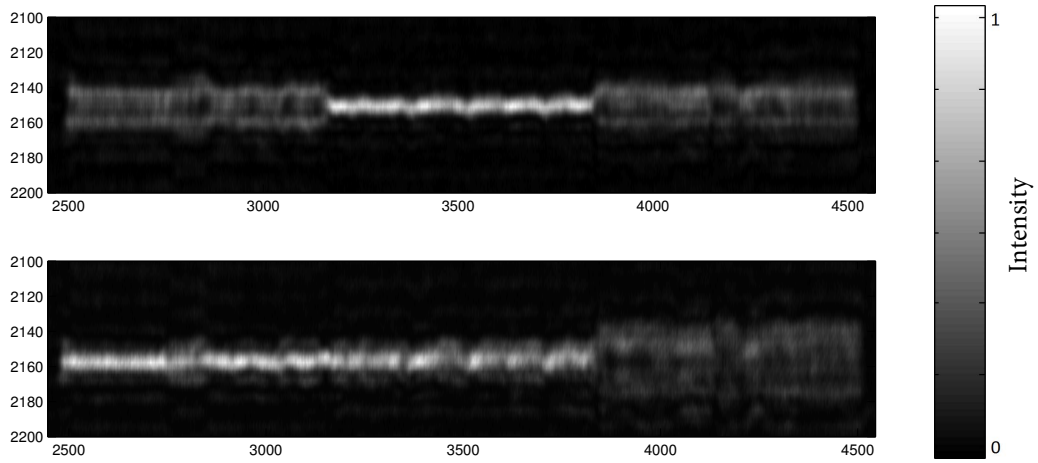


Figure 6.25: Single line on 3 constraint planes separated by $758\mu m$. Top shows middle constraint in focus, bottom shows an intermediate focal image between two constraint planes.

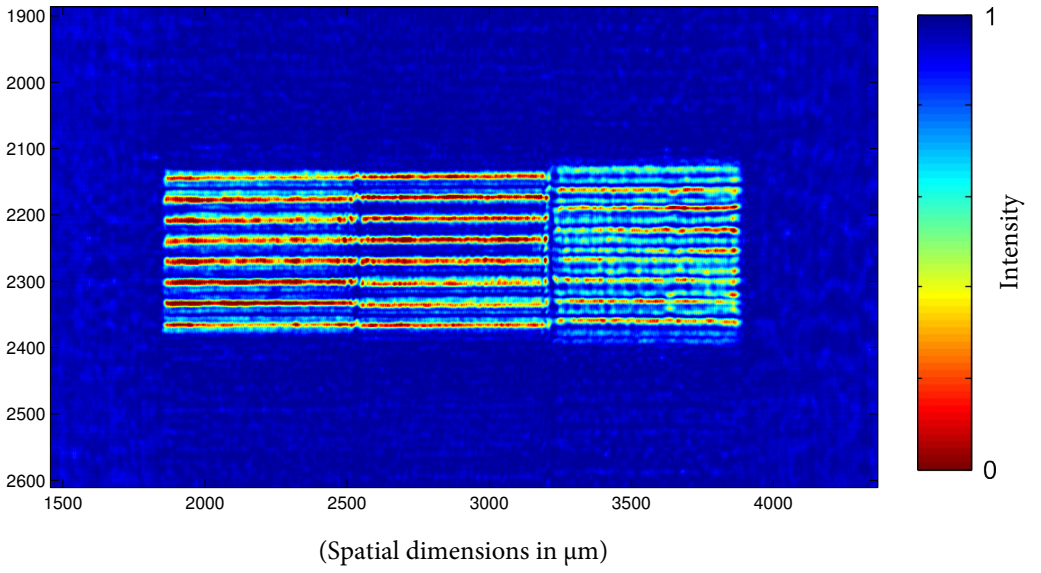


Figure 6.26: Bus pattern with inter-constraint distance set to $758\mu m$.

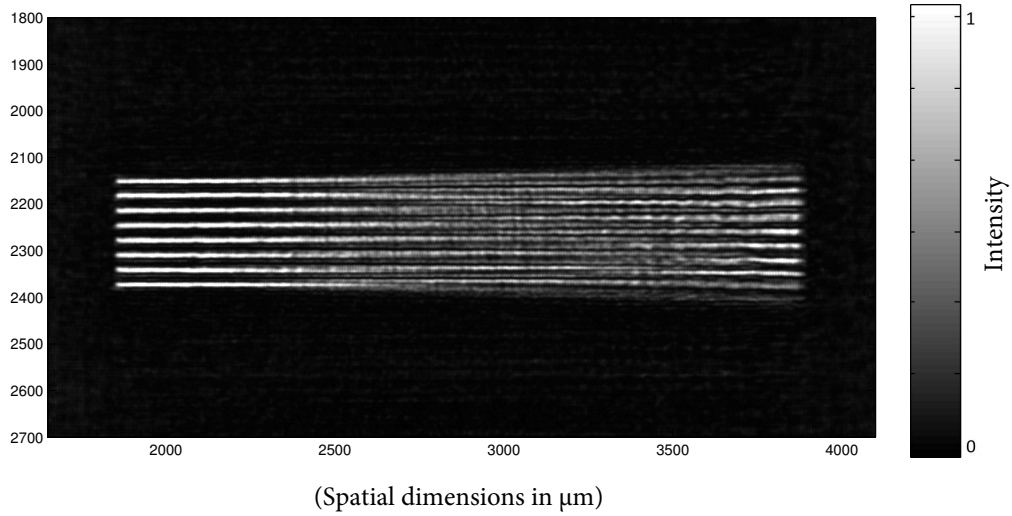


Figure 6.27: Bus pattern with inter-constraint distance set to $100\mu m$.

6.8 Conclusions

In this chapter we have implemented and explored a multi-plane iteration scheme. We have demonstrated the issues with this basic method such as proximity constraint plane interference, and incorporated fixes which allow us to pattern a three dimensional surface under reasonable constraints. It has been shown how considering and suppressing the freely propagating energy in our iterative simulation can prevent gain issues which would hinder generation of the intended pattern. We have gone on to investigate the required constraint plane separation to form a continuous lithographic exposure by considering a minimum contrast for viable exposure.

By comparing how the contrast is affected in three different cases for a simple set of patterns (two zone plate points, two iterated line sections and finally one continuous iterated tilted line pattern) we have established a useful lower bound for the interplane constraint separation. Though the values derived are quite specific to the simulation set up and patterns, it is hoped that the ratio of system DOF to constraint separation distance will provide a useful baseline for future experiments. The core novel result here, however, is that we have demonstrated that it is possible to form a suitable intensity pattern for these line patterns using this multi-plane partial constraint method.

The lines patterns created by this method are not constrained by the width of the hologram pattern as is the case for simple analytical approximations and can be made arbitrarily dense up to the diffraction limit. The quality of the result does depend on the complexity of the constraint pattern on the surface as well as the complexity of the pattern.

In the following chapter we validate this by experiment by showing a bus of lines in a sloped exposure many times the DOF of the optical system. Example practical measurements using our SLM based phase modulator system and captured on CMOS sensor have verified that applying our iterative algorithm along with the modifications discussed in the previous chapter result in patterns which, whilst noisier than those shown in simulation, are still valid high contrast patterns.

Chapter 7

Application Focused Examples

7.1 Introduction

Having developed a viable patterning method in 3D and established the basis of its limitations, we move now onto finding and demonstrating some useful application-led examples. Alongside this, some new methods are developed to help us cope with yet more extreme substrates and their effect on our iteratively generated holograms is examined.

Three examples are presented in this chapter. The first shows a dense bus of lines over a sloped "dog-leg" substrate. This demonstrates the power of this methodology to maintain a sensitive high resolution pattern over a slope. This kind of interconnection pattern at a slightly different scale might form a useful interconnection for electronics packaging, or anisotropically etched silicon edges.

Secondly, an antenna pattern is presented. Most small modern antennas are planar in design due to manufacture constraints. Building micro-scale antennas on 3D substrates has the potential to augment their properties in new and novel ways. In our particular case, the antenna proves to be an interesting test for active modulation averaging techniques, allowing for larger areas to be exposed on a 3D substrate.

Thirdly, we present a method in which we combine our holograms with a grating which allows for high resolution patterns to be exposed on extremely steep substrates. In this case the application is electronics interconnection over the edge of a vertical silicon die i.e. a potential replacement to wire-bonding methodologies.

7.2 Dense Sloped Bus

Research carried out at Durham and Sheffield universities has previously demonstrated high resolution line interconnections on a sloped surface. However, the limitations of the analytical holograms used are particularly prevalent when attempting to pattern a dense bus of interconnections.

Here we present an example of a dense bus pattern projected onto a "dog-leg" slope surface. The hologram is optimised using an iterative multi-plane constraint algorithm which has been seeded with an analytical approximation.

This example serves as a basis for demonstrating how thin lines patterns can be applied in 3D and serves as an experimental verification proving that 3D lithography with such a multi-plane constraint pattern is possible.

Furthermore we once again assess the inter-plane separation required to form this more complex dense bus pattern.

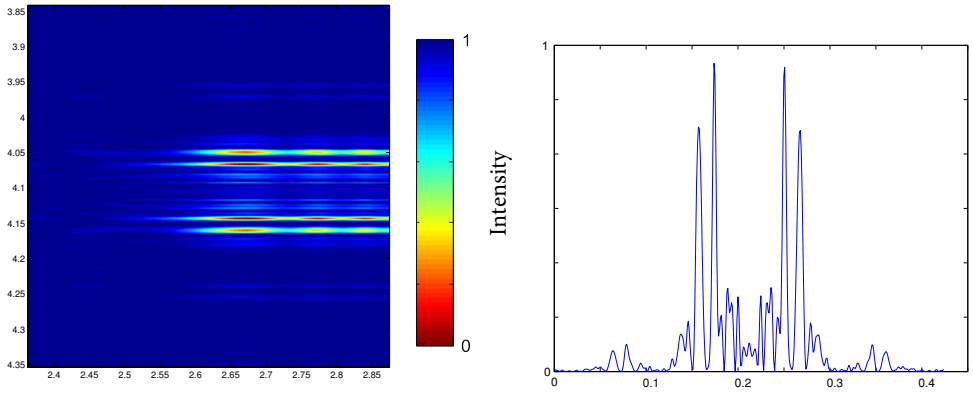
7.2.1 Analytical Approximation

The size of an analytical hologram fundamentally limits the pitch of the lines producible. In an analytical case, holograms should be computed such that they do not overlap for the lines patterns produced resulting in a limited line pitch. For a given feature size the hologram width can be calculated by the grating equation and the focal distance.

$$\text{Width} = z \tan\left(\sin^{-1}\left(\frac{\lambda}{d}\right)\right) \tag{7.1}$$

Where z is line focal depth and d is once again feature size. When increasing the density of the lines beyond this limit to the extent that the holograms overlap, interference between each separate hologram starts to cause inconsistencies between lines patterns. An example of this is shown in Figure 7.1.

The relative phase of the overlapping line holograms can be modulated to minimise problematic cross-interference and produce a more uniform intensity distribution. This issue can become more complex in 3D as angled fringes are likely to overlap and cause yet more complex interaction patterns. Moreover, brute force optimisation of the relative phase of each of the line holograms in this case is computationally unachievable because of the number of possible combinations. Even assuming a relatively coarse quantised phase optimisation with each line hologram having eight possible phases, the number of combinations is above 16 million. For a field of the size suggested on a low to average GPU such as the the Tesla 1060c we would expect the time required to composite, simulate and analyse a single phase



(a) (Dimensions in mm) (b)

Figure 7.1: Overlapping line holograms, (a) simulated intensity pattern, (b) intensity profile.

pattern to be on the order of ten seconds. This leads to a total expected computation time of around five years.

Figure 7.1 shows a set of planar lines placed at the average focal distance of the desired 3D pattern, generated by overlapping and quantised phase holograms. The parameters used in this simulation are comparable to those shown in table 7.1. Even in 2D it is clear that this overlap approximation and phase-only constraint is ineffective, and that it would be very difficult to produce multiple lines using such an approximation.

7.2.2 Iterative Example

To overcome this density limitation and gain some of the other advantages seen in our previous examples we turn to our iterative method. We know that we can generate arbitrarily dense patterns but that the noise in the final image can become the limiting factor.

The parameters used for this dense bus pattern are shown in table 7.1

The depth of focus at the front of the image volume for these parameters is calculated to be $1.236mm$ by Equation 6.1.

We then test this pattern with varying numbers of constraint planes (2 to 128) in simulation after 50 optimisation iterations. Contrast in this case is measured by sampling intensity from a quantised surface at the pixel pitch of the hologram. Surface normal effects are ignored for this measurement.

Figure 7.2 shows the intended pattern structure and 3D surface, and the corresponding automatically-generated constraint patterns for the flat-slope-flat surface

Hologram-Image Separation (z)	Wavelength (λ)	Sample/Pixel Pitch
$16cm$	$405nm$	$4\mu m$
Image Width/Height	**Simulation Width/Height**	**Feature Width**
2048 pixels	4096 pixels	2 pixels
Constraint Low Value	**Constraint High Value**	**Number of Iterations**
0.1	1.0	50
Line Pitch	**Slope Depth**	
$24\mu m$	$4096\mu m$	

Table 7.1: Simulation and image parameters for dense bus test.

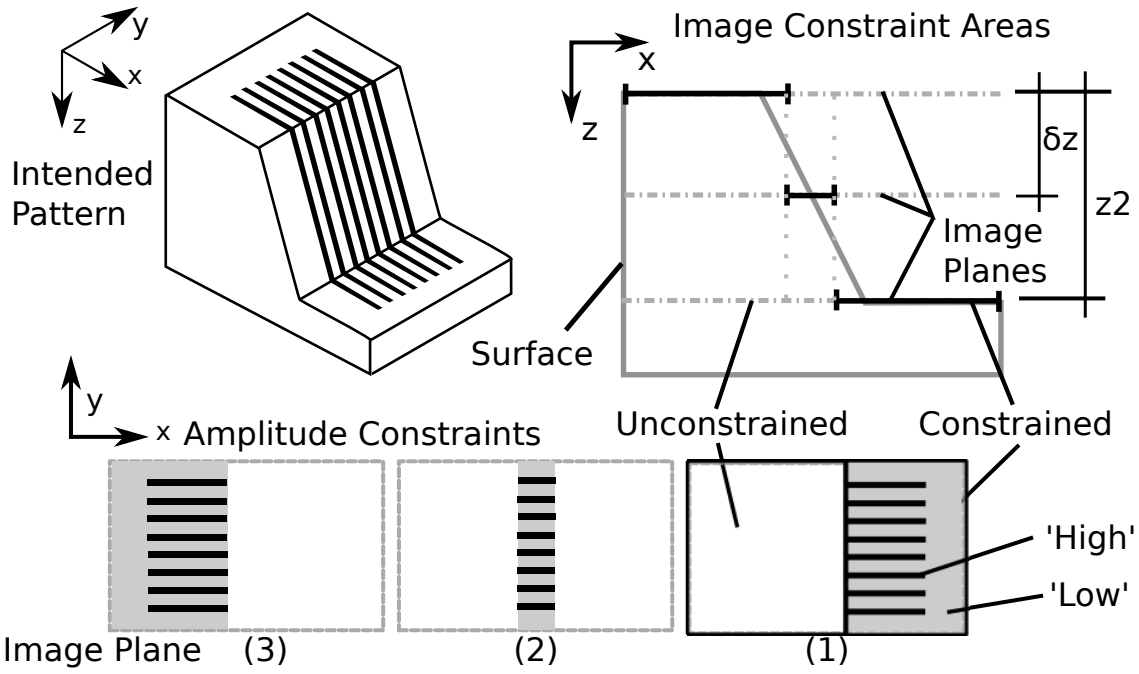


Figure 7.2: Intended pattern and example planar breakdown.

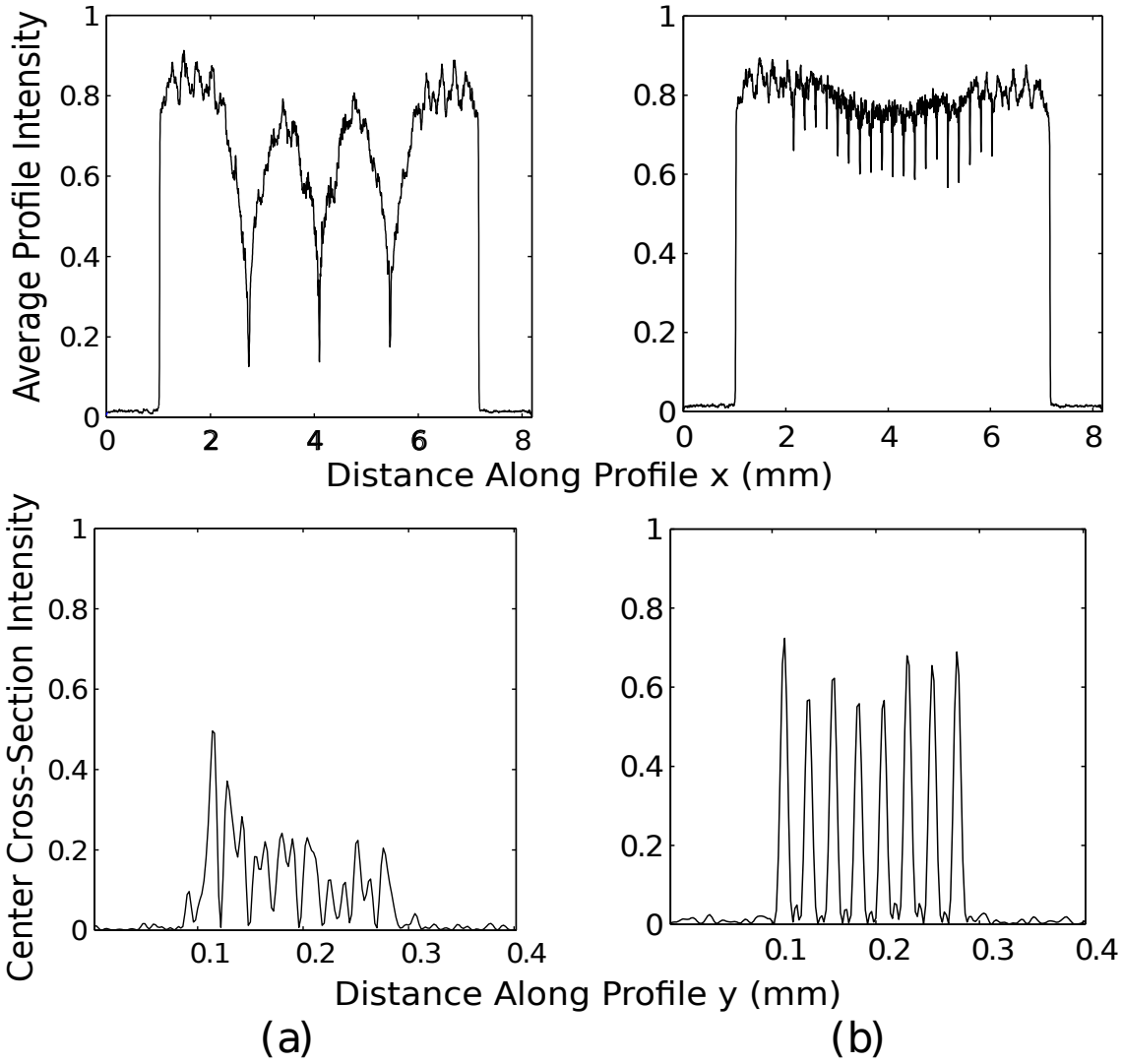


Figure 7.3: Simulated surface profiles for different inter-plane separation distances. (a) 1365 μm and (b) 215 μm

(surface and patterns not to scale).

Our standard SLM exposure system shown in Section 4.7 is used with an SLM to focal pattern distance of $16cm$. The optical bandwidth of the SLM at this focal distance is the limiting factor behind this decision. At this distance, filtering of the transfer function is not required. This simplifies some assumptions made in the algorithm. The system is sampled and analysed at $4\mu m$ sample pitch. Iterations are seeded with an analytical approximation of the line pattern made from superimposed cylindrical line holograms.

Two pairs of example surface profiles are shown in figure 7.3. (a) and (b) show profiles along (top) and across (bottom) the line bus pattern in the centre of the intended image on the focal surface. In this case the focal pattern amplitude has not been adjusted for surface angle.

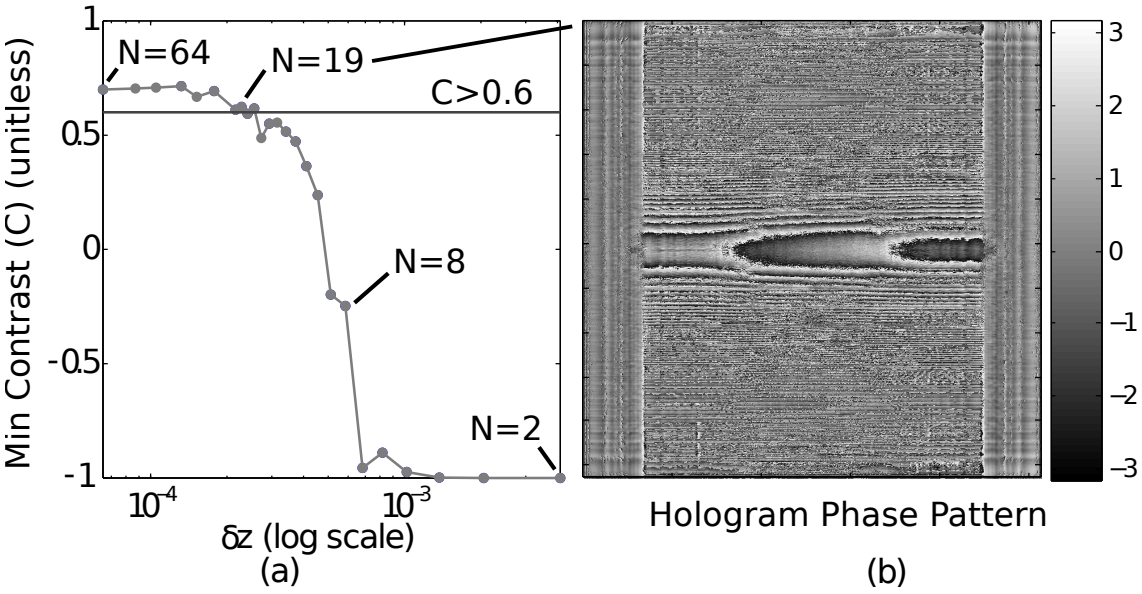


Figure 7.4: Simulated surface contrast measurement and example phase hologram pattern

Figure 7.4 shows contrast for the bus pattern for varying inter-plane separation distances. When comparing this to our calculated depth of focus we find that our contrast metric reaches a suitable level of greater than 0.6 when the inter-plane separation is less than $228\mu m$ or 0.18 times the calculated depth of focus. This is significantly smaller than the measured required separation for a single line with similar parameters, but this increase is not unexpected as the complexity of the pattern (number of lines and variation in substrate shape) has increased.

Figure 7.5 shows two glass slides exposed with different sections of the pattern. For experimental exposure, the hologram was implemented on an $8\mu m$ sample pitch SLM ('Pluto' from Holoeye Photonics AG) by quantizing and re-sampling the phase pattern. This device is illuminated by an on-axis expanded laser beam (Coherent 'Cube' $405nm$ $50mW$). The photoresist used was BPRS200 from Fuji Photo Film Co Ltd. This layer was approximately $2\mu m$ thick. (a) shows a flat section at the top of the slope and the line termination. (b) shows a tilted section exposed onto a tilted glass slide. Both show good agreement in feature size, and regularity. The arrows on figure (b) indicate sections where we would expect to see noise from the changing of partial constraints, and indeed some slight variation is observed. This demonstrates successful practical optical lithographic exposure of a 3D focal pattern.

For comparison, we have included experimental focal images of the image intensity pattern with the top region and the middle region of the slope in focus 7.6. In the sloped pattern it is clear that the image is quickly destroyed in out-of-focus regions. This proves that a single image plane would not be viable.

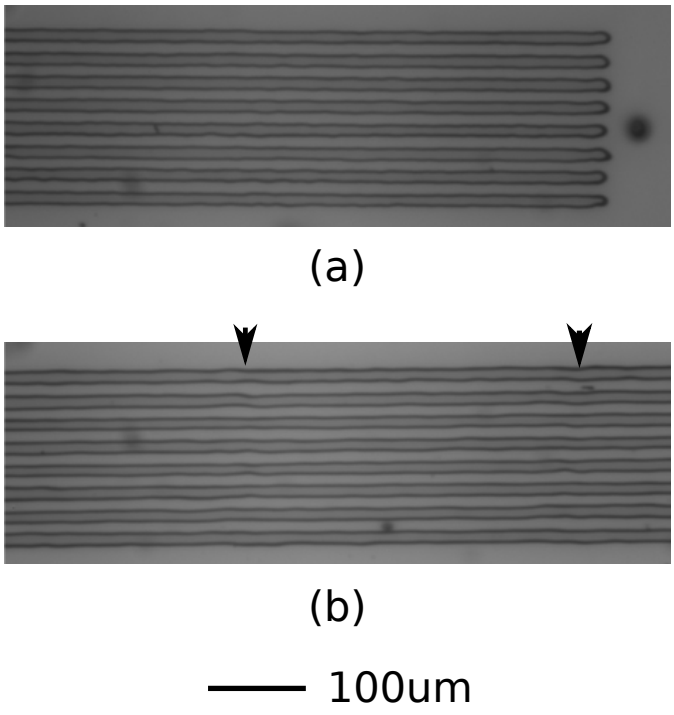


Figure 7.5: Planar and tilted bus sections, experimentally exposed and developed on a glass substrate.

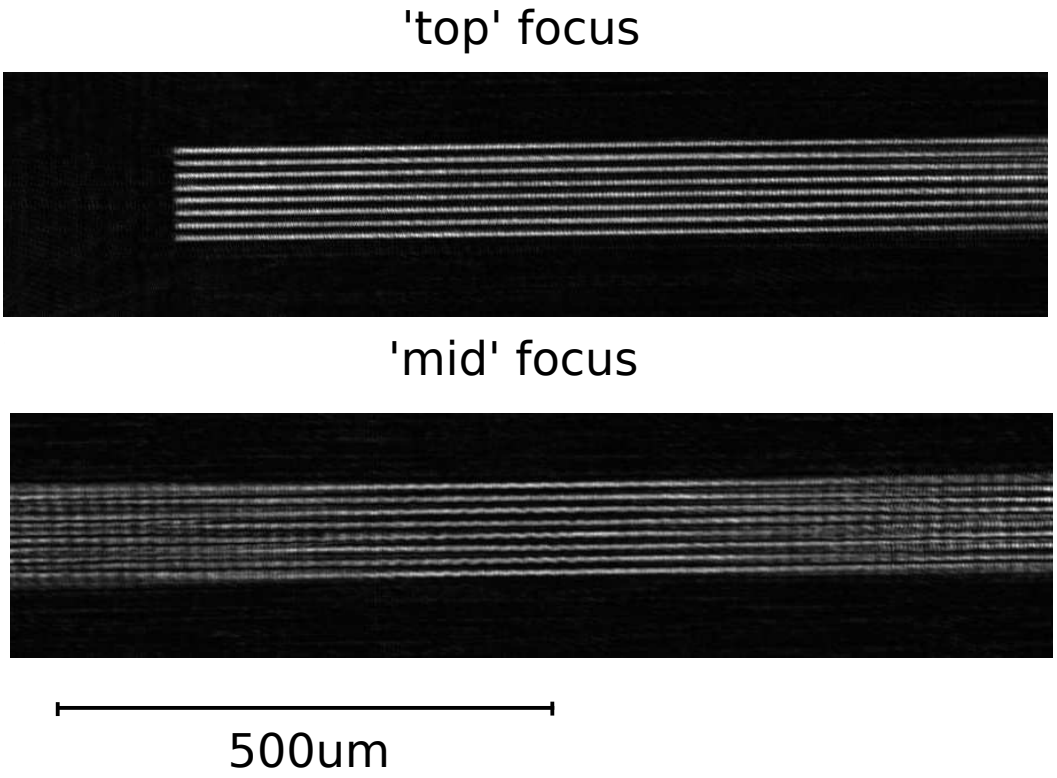


Figure 7.6: Experimentally recorded intensity images cross sections at top and centre of the focal image.

7.3 3D Antenna

7.3.1 Introduction

In previous work related to the development of analytical approaches to holographic lithography [26] the research was able to generate a spiral antenna on a centimetre scale cone. This antenna design was chosen to show the potential benefit of 3D lithography, and by taking a spiral design out of plane it is noted that there are potential improvements in the properties of the radiation pattern and antenna bandwidth.

Following up this antenna design work a new antenna pattern was designed in conjuction with Sheffield University (Dr Gavin Williams, Dr Jesus Toriz Garcia, Dr Luke Seed and Prof Peter Ivey) [14]. The details of this patterns proposed improvements have not been a focus of this work and will not be discussed here. The design, however, is interesting as a test for the purposes of our more general approach to 3D lithography, as it could not be achieved with traditional lithographic methods or analytical holographic methods under phase-only constraint.

7.3.2 Pattern

The substrate is a 1cm diameter hemisphere, with a set of eight meandering lines covering its surface. Our modulator is not of sufficient size to expose the whole substrate in a single illumination, therefore we developed a single meander pattern and optimised it for a tilted hemisphere. Figure 7.7 shows the intended pattern on a 3D representation of the surface and figure 7.8 shows this pattern projected onto a flat 45° inclined plane.

This pattern is geometrically projected onto a flat plane at a 45 degree angle to the hemisphere to create the plan image pattern 7.8. The 45 degree angle minimises the depth of the substrate (in the direction of illumination) and ensures that as little resolution as possible is lost in the projection to a uniformly sampled plane in simulation. With this method, the angle between the most extreme parts of the surface and the illumination direction is minimised.

7.3.3 2D Experimental test

Initially a flat version of the antenna pattern was created to examine implementation issues. In this case the experimental parameters were set up as in table 7.2

In the initial case the 'filtering' method was used to limit the field feature size in the hologram to approximately $8\mu m$ i.e. a two-fold decrease from the simulation

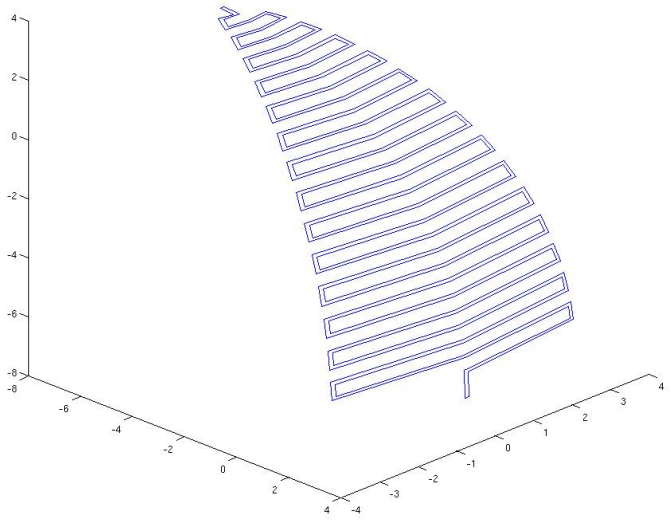


Figure 7.7: Antenna outline pattern

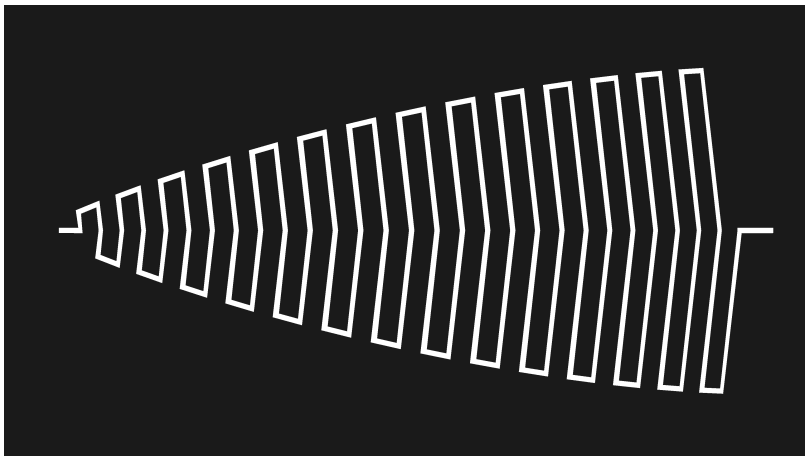


Figure 7.8: Pattern projected onto 45° plane.

Propagation distance (z)	Wavelength (λ)	Sample/Pixel Pitch
$8cm$	$405nm$	$4\mu m$
Image Width/Height	**Simulation Width/Height**	**Hologram Size**
3840 by 2160 pixels	7689 by 4320 pixels	3840 by 2160 pixels
Constraint Low Value	**Constraint High Value**	**Number of Iterations**
0.025	1.0	100

Table 7.2: Parameters used in planar antenna iteration experiments.

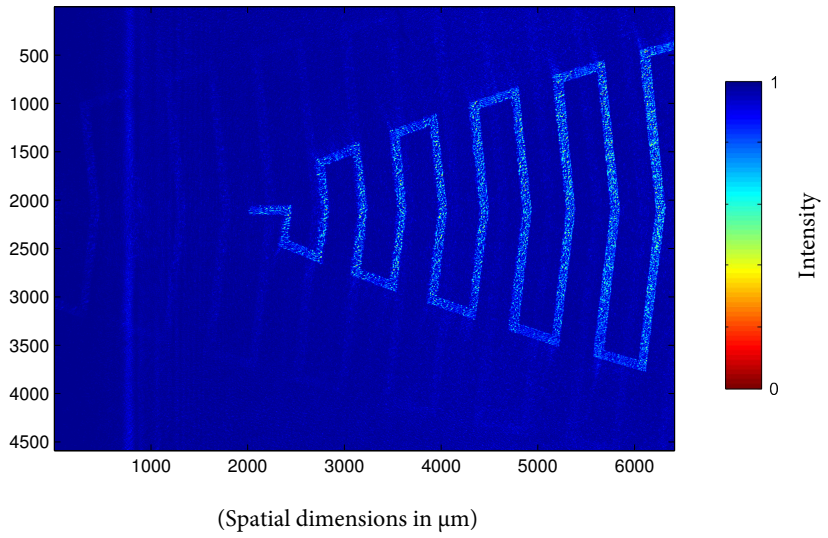


Figure 7.9: Experimental reconstruction of iterated antenna pattern.

pixel size, as minimum feature size is not an issue with this pattern. Furthermore we once more match a maximum feature size of $8\mu m$ with the implementable $8\mu m$ pixels of our SLM such that minimal issues occur from re-sampling.

Figure 7.9 shows an experimental reconstruction of the planar pattern optimised with this iteration method. We note once again that the pattern is well defined. An average contrast (note: not minimum as used in the previous two chapters) of 0.67 is measured between profiles in the high and low regions. We do, however, also note the visible first order image which overlaps the pattern in this section due to the scale of the image and its proximity to the SLM.

Figure 7.9 shows an experimental reconstruction of the same image, with both the filter and pixelation methods of bandwidth control and implementation simulation applied. The pixelation method in this case resamples the hologram after its phase-only constraint and sets each two-by-two pixel group to a single value.

In this experimental image we observe an increase in average 'high' region intensity of 70.1% resulting in an improved average contrast of 0.78. Furthermore, by taking profiles of the first order images outside of the image region, we observe a 20% reduction in the average intensity of the first order image when applying this pixelation approximation in conjunction with the filtered iteration. We believe that the pixelation routine causes these improvements by being a more accurate model of the limitations and implementation of the re-sampled implemented hologram. By 'pixelating' the hologram the first order image is also observed in the simulation giving the algorithm the opportunity to compensate for its presence.

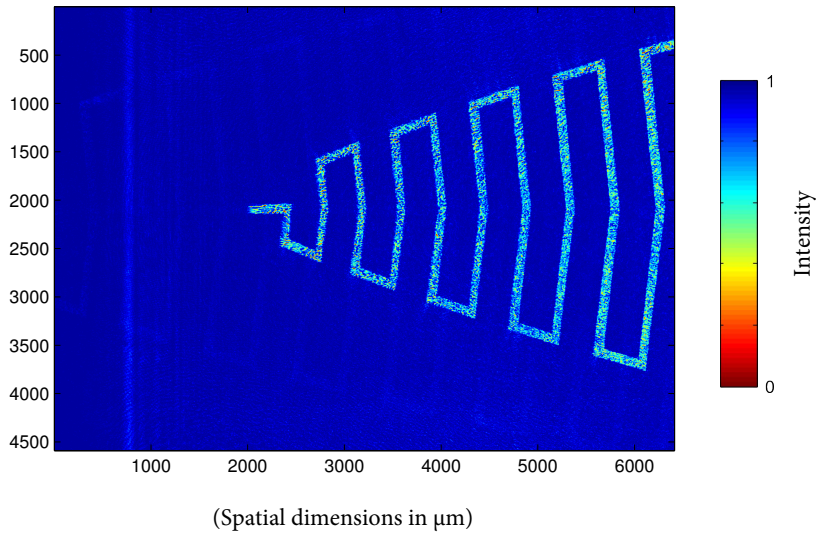


Figure 7.10: Experimental reconstruction of iterated antenna pattern designed with filter and pixelation applied.

One final feature observed in both of these experimental tests is a bright oscillating band of intensity at the left edge of the image frame. This band is not caused by the interface between the image region or implemented hologram at the edge of the SLM. It is in fact caused by a small, unmodulating reflective region at the edge of the SLM. Placing a small window aperture close to the SLM to block out this reflective region was found effective in experimental exposure tests.

7.3.4 3D Pattern

Because the depth of the substrate is minimised by tilting the hemisphere to expose the meander pattern, the number of required simulation planes is also minimised. In this case, the pattern has a depth of approximately $5.27mm$. The depth of focus of the optical system at its top resolution is once again $1.263mm$ (limited by the filter applied to the propagation kernel at 0.25 times max bandwidth). From our single line example we observed a consistently high contrast from an inter-plane separation of approximately 0.5 times the DOF of the optical system or smaller. In the previous bus of lines example we observed that a more stringent 0.18 times the DOF was required. However, in this case, when we average many holograms together and noise is removed from the image, the feature size is much larger than the minimum feature size of the system. It would be easy to expect the average DOF of the image to also increase, potentially reducing the requirement for the number of constraint planes.

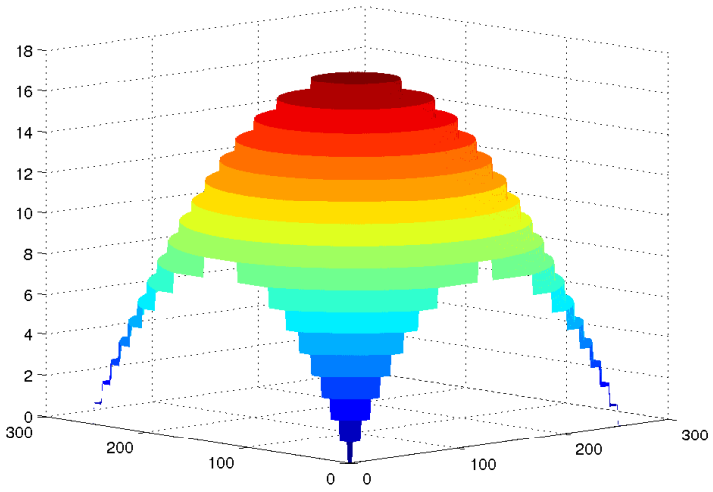


Figure 7.11: Quantised spherical surface.

We choose here to show an example which is 0.25 times the DOF of the system. In this case that equates to dividing the pattern into at least 18 constraint planes. By our previous measurements this should be approximately appropriate to begin to form continuous patterns (i.e. it is between the two previous measurements of 0.5 and 0.18 times DOF).

Under the multi-plane iteration algorithm and the same implementation constraints as described in the flat image experiment with a random phase seed pattern the iterative algorithm results in hologram as shown in figure 7.12.

The features of the hologram are once again interesting to assess. Despite the random phase pattern, a clear delineation in the structure of the phase pattern is observable and follows the shape of the underlying antenna pattern. This delineation once again related to the maximum diffraction angle (i.e. filtered bandwidth) of the iteration and simulation which limits the useful extent of the hologram.

An experimentally recorded and averaged intensity pattern from 20 holograms is shown in figure 7.13. We note here that the pattern itself exhibits very little variation in feature size or quality outside of the focal region. The centre rightmost section of the image is known to be in the topmost region of the quantised constraint. The leftmost part of the pattern is in optimal focus, achieved by moving the CMOS sensor to maximise the pattern definition in that constraint region. In this case it is clear that the pattern DOF is likely more than sufficient to expose the whole pattern with only one flat focal image. However it is interesting to note that the constraint regions are more sensitive to the focus depth of the constraint pattern, than would be implied by the depth of focus related to a larger feature size.

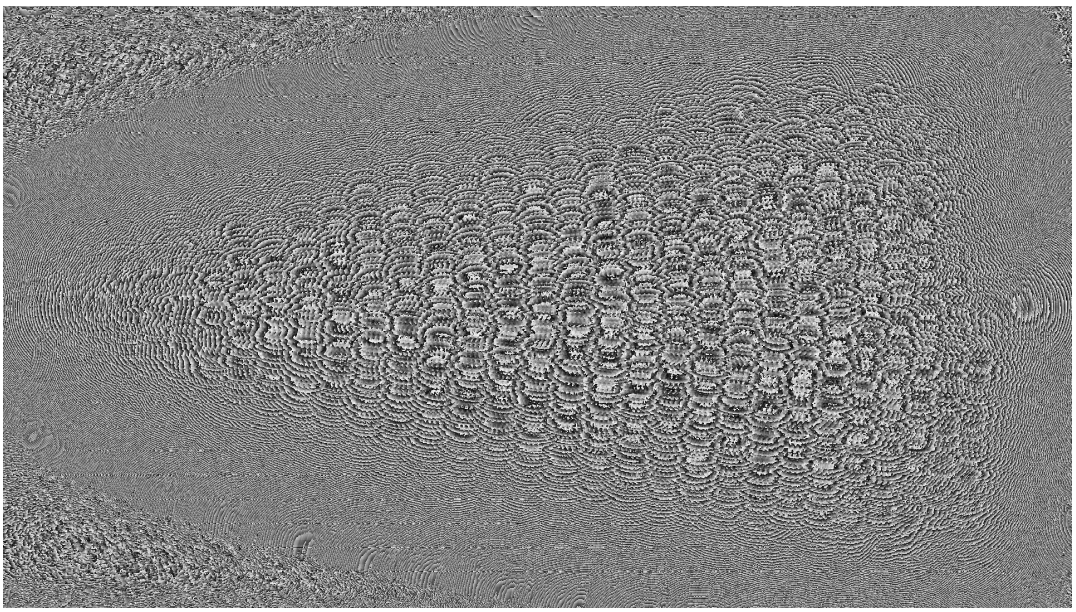


Figure 7.12: Iterated antenna hologram phase pattern.

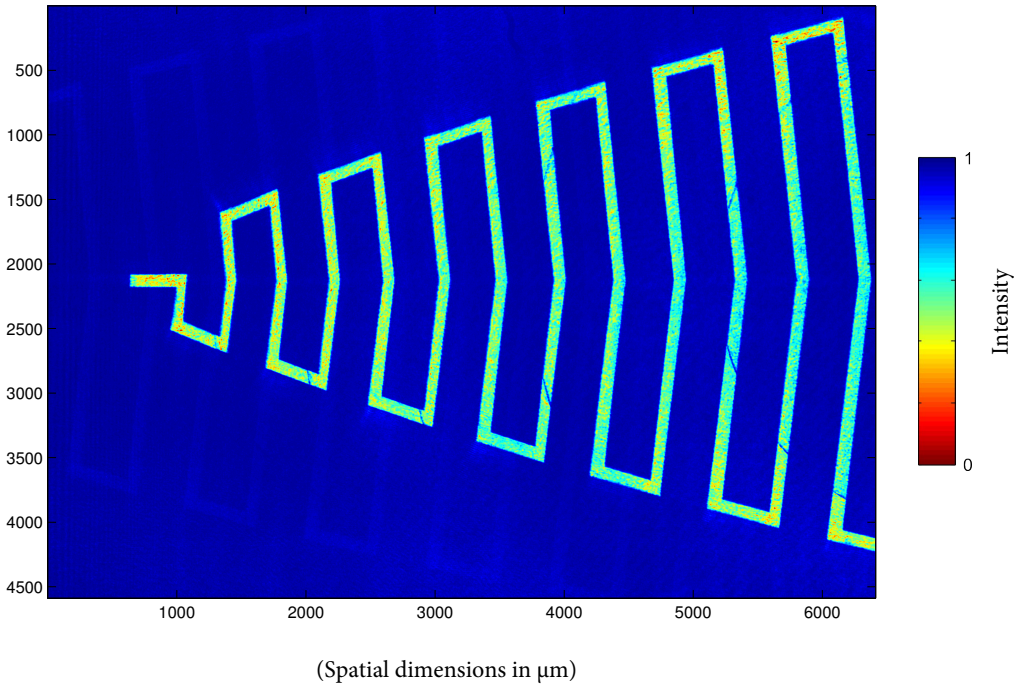
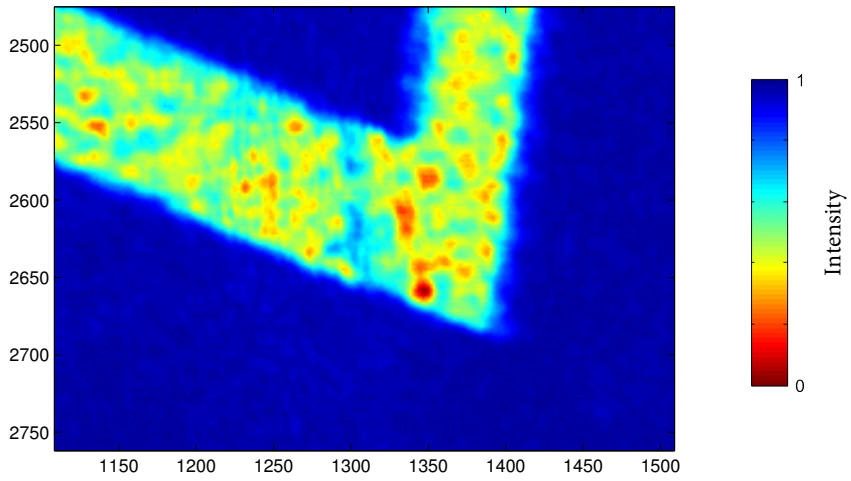
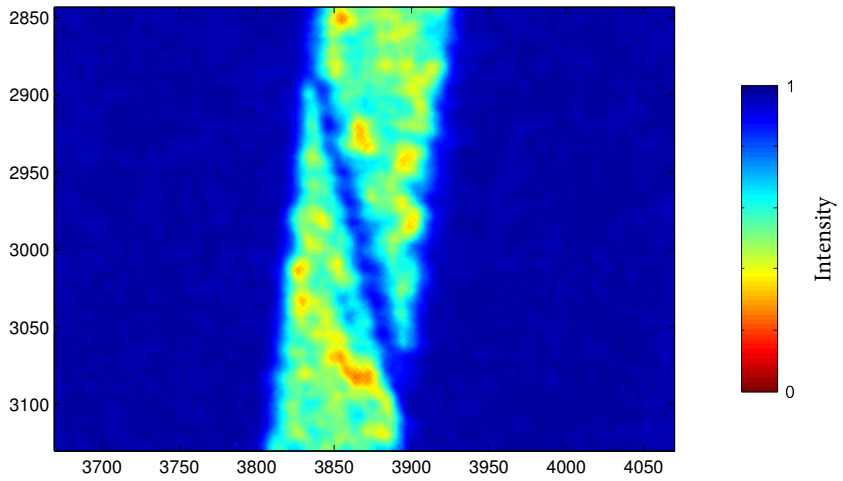


Figure 7.13: Experimentally recorded averaged 3D focal image subsection. Antenna 'tip' section in focus. Dimensions in μm



(Spatial dimensions in μm)

(a)



(Spatial dimensions in μm)

(b)

Figure 7.14: (a) Antenna boundary region intensity pattern in focus. (b) A separate out of focus boundary region.

Figure 7.14 shows examples of average intensity pattern sub regions around partial constraint boundaries. The in-focus region barely registers a discontinuity, whereas a similar boundary in an out of focus region shows a clear line of demarcation between the two constraint regions. The phase pattern on this boundary has been unable to optimise for a continuous intensity pattern and has instead formed a high resolution high contrast boundary between two high intensity regions. This boundary is likely formed similarly to a phase contrast mask feature, a phase jump over the boundary leading to a null in intensity value. Holograms with such phase jumps could potentially be manipulated to form these kinds of features if required.

Given the much larger expected depth of focus expected from the larger feature size this sensitivity is of some surprise. Since the pattern definition in non-boundary regions remains well defined we surmise that in this case splitting the pattern into many constraint regions is having only a negative effect on the pattern quality for this larger feature size at this relatively small pattern depth.

This is not to say however that this method is not viable. Though the advantages are not obvious in this specific case a similar thinner pattern is likely to show greater improvement.

Figure 7.15 shows the antenna pattern exposed and developed on a glass hemisphere. For the practical exposure case 128 constraint planes were used which is a large oversampling in depth of focus to ensure highest pattern quality.

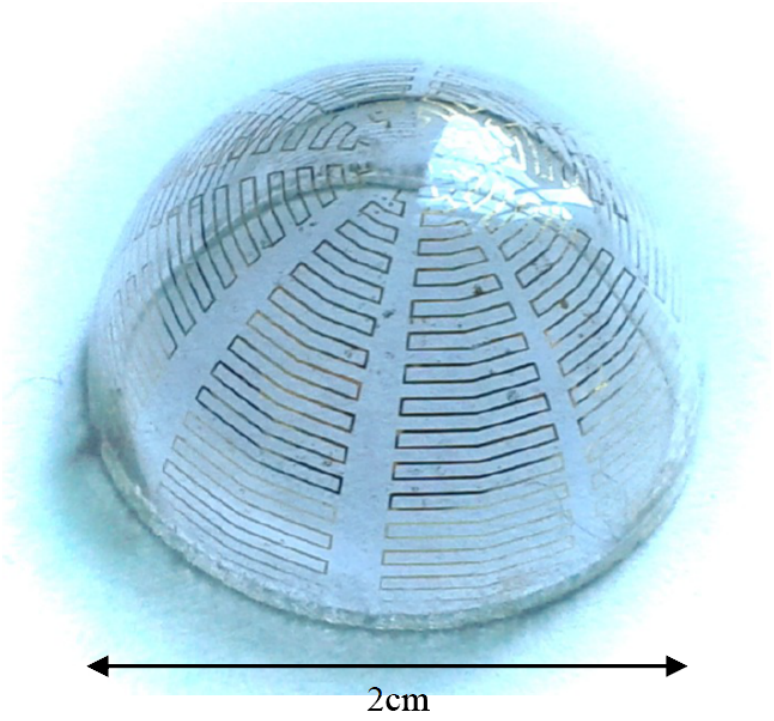


Figure 7.15: Antenna pattern exposed multiple times and developed on glass hemisphere 2cm diameter.

7.4 Vertical Patterning

7.4.1 Introduction

In considering our examples it is clear that substrate geometry is probably the most severe limitation on the method. As substrate angle increases, intensity is decreased by surface obliquity factor and also increased reflection. At more extreme angles the surface itself eventually occludes parts of the hologram.

The case of a completely orthogonal surface to the beam propagation direction is of most interest to us. Not only does it poses a significant challenge to 3D lithography, there are also several potentially useful applications which could be approached if this limitation were broached.

Antenna manufacturers Sarantel [84] previously produced a line of helical antennas patterned onto the outside of a cylinder. Their current process involves a rotating substrate and sequential lithography. Being able to pattern at right angles to the beam direction would allow this process and similar vertical patterning problems to be carried out in one exposure in a potentially faster or simpler way without the need for serial techniques.

Interchip connections are another area that could potentially benefit. Currently most interchip connections are formed by wire bonding or TSV. By being able to

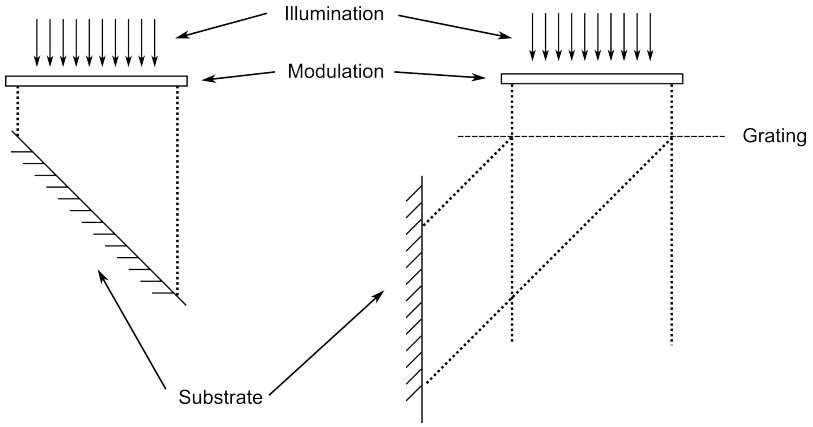


Figure 7.16: Diagram of grating exposure system.

pattern the side of a silicon chip, we could pattern these connections lithographically without need of special one off processes.

7.4.2 Approach

By a combination of an optical grating and an analytical hologram [11] our group at Durham and Sheffield have shown that such extreme geometries can be patterned with only minor modifications to our exposure system.

Nellissen's technique [85, 86] of using a grating to alter the direction of incident light on a mask is potentially significantly improved by considering more complex diffractive masks as higher resolution and more complex patterns become achievable.

If the grating used to augment a hologram were a perfect blazed grating the majority of the energy of the projection would be shifted into the +1 order image. This process is illustrated in 7.16

The first order image is further distorted by this process for which a compensation is required in the hologram design. Generally a multiplication by a blazed grating results in a shift of the spectrum, or equivalently a rotation of the ewald sphere around its axis. This is a fairly complex rotate and interpolate function to apply to the spectrum for image simulation. It may also require a significant increase in simulation sample pitch as discussed by section 2.6.

Thankfully, our group at Durham and Sheffield have found that the distortion can be geometrically compensated for by considering a fixed focal distance for a given point from the grating modulation plane, and rotating this vector around its corresponding focal point on the grating plane.

The result of this transformation is that a sloped surface focal image is rotated by the grating angle, and compressed in one dimension by a factor given by

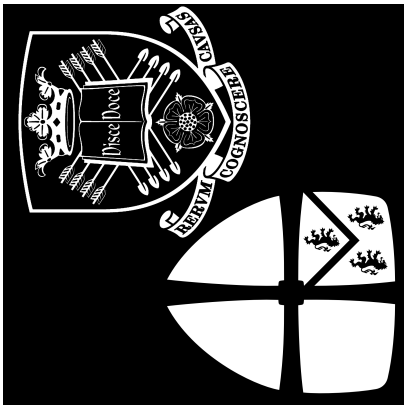


Figure 7.17: Crests test pattern.

$$\Delta W = \cot\theta \tag{7.2}$$

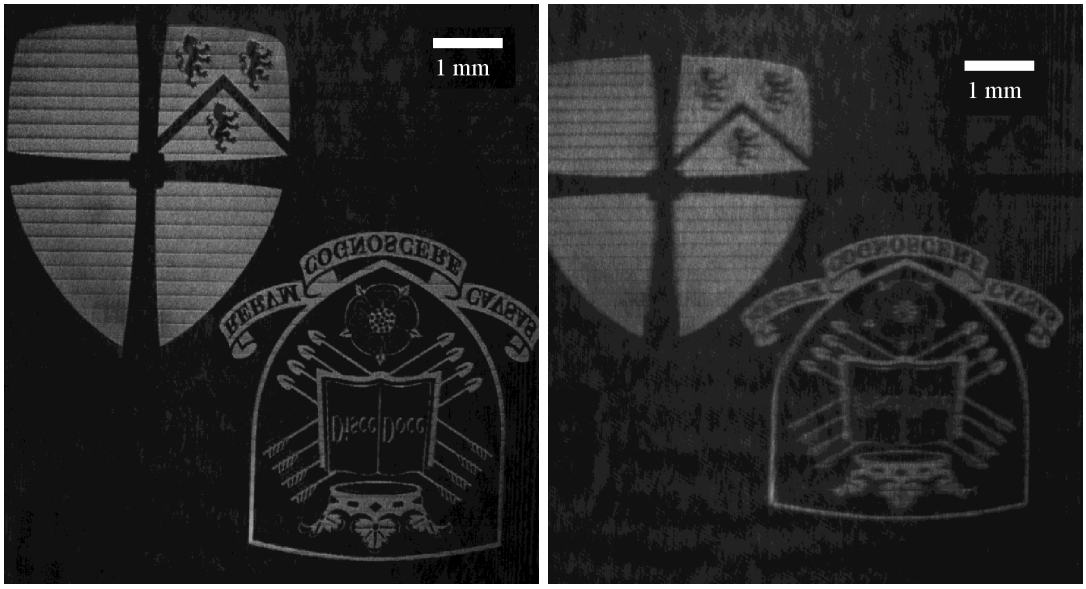
Given a grating of approximately 100 lines per millimeter and a wavelength of $405nm$ we have a first order angle of approximately 23.9 degrees. Therefore an image on a slope of 66.1 degrees will result in an image which is perpendicular to the grating pattern.

7.4.3 Test Patterns

Figure 7.17 shows a pattern containing the Durham and Sheffield crests which we have used as a complex test. A sloped focal surface has been formed by dividing the image pattern into N quantised partial constraints. The iterative multi-plane algorithm is then used to optimise 40 holograms. Those 40 holograms are cycled on an SLM modulator during exposure to create a high resolution averaged pattern. A test sloped intensity pattern taken with the camera oriented to the correct sloped focal plane is shown in figure 7.18 (a). A camera image showing vertical image of the first order pattern is shown in figure 7.18 (b).

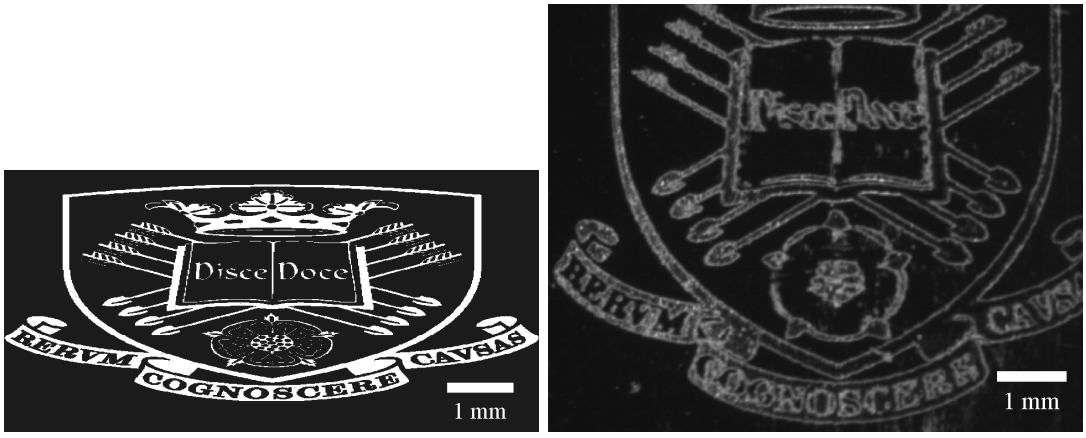
In this case the constraint plane separation was set to approximately 0.15 times the DOF of the optical system to ensure that the pattern would be suitable. Despite this, some banding in the image is clear. Furthermore, it is clear that some detail has been lost in the first order image. This is likely due to the quality of the grating used degrading the image rather than any other physical effect.

Finally, we present a scaled Sheffield crest (figure 7.19 (a)) and its pattern exposed on a glass substrate (figure 7.19 (b)) performed by Dr Jesus Toriz-Garcia. The glass slide is orientated to catch the vertical focal image. Surprisingly little detail is lost in the first order as it can be seen that the fletching on the arrows in



(a) (b)

Figure 7.18: (a) Sloped zero order and (b) vertical first order focal images projected by grating.



(a) (b)

Figure 7.19: (a) Scaled Sheffield crest and (b) vertical exposure.

the crest is visible.

Chapter 8

Summary and Conclusion

8.1 Summary and Discussions

The aim of this thesis was to develop an optical design methodology based around iterative optimisation algorithms commonly used in research areas such as holographic modulation for display. This methodology allowed us a more general construction of image patterns for lithographic purposes than had been previously achieved using line holograms [26].

In chapter 2 we began by looking at some of the fundamental methods for optical simulation and comparing their accuracy, suitability and computational efficiency. From this we found that the AS propagation method was suitable and appropriate for execution on modern hardware at the kind of scales we require for simulation. We then looked at some simple examples of phase-only holograms constructed from cylindrical Fresnel lenses and showed that under certain conditions (such as bringing those line patterns together into denser bus structures) the production of such phase-only patterns was not suitable.

We limited ourselves to phase-only patterns for a number of reasons. From a lithographic point of view phase-only patterns allow the image to have a higher diffraction efficiency by not removing any energy from the field. From a practical point of view, the concept of a phase-only constraint matches closely with useful existing optimisation algorithms based on the GS approach, and also matches closely with what can be achieved in hardware, i.e. phase-only SLM modulation.

In chapter 3 the structure and merits of some iterative optimisation algorithms are discussed. Lithographic optimisation methods are discussed before moving on to methods derived from the Gerchberg-Saxton algorithm, which is often the basis of phase-only hologram design or holographic phase retrieval. Variants of this algorithm have been used extensively to design phase hologram patterns in display,

beam shaping and holographic optical tweezers, as well as lithography.

After discussing and comparing parts of the optical set-up required to project a holographic phase-only pattern in chapter 4, we settled on a useful optical set-up based on collimated laser illumination on an SLM. However it is interesting to note that potential alternatives closer to existing mask exposure systems could be used if coherence were sufficiently increased.

In chapter 5 simple GS implementation was tested, which was found lacking due to a collected noise spot and a very noisy image pattern. We then moved on to describe and implement a basic planar iterative scheme using the AS transfer function.

In a set of the most fundamental cases we assessed the suitability of each pattern for lithography using our minimum contrast metric, which was more helpful than SSE in assessing the viability of the resulting patterns for lithography. In these patterns we found that thin line features were producible, but that larger areas were plagued by 'speckle-like' noise which could not be optimised away by the iterative algorithm. Furthermore, we discovered that the choice of initial phase pattern used to seed the algorithm could have a significant effect on the resultant image, and that choosing a plane phase or analytical pattern for our thin feature tests was often appropriate, resulting in a higher contrast.

Moving on the limits of our proposed planar algorithm were assessed. We started with the propagation distance limit of the transfer function. By a method analogous to that of Sypek in [20], we derived a limit on the propagation distance of the AS transfer function and implemented a similar step and window process to achieve longer propagation distances. This did result in some slightly erroneous behaviour when examining the error reduction of the algorithm, however it was evident that the resultant patterns were relatively unaffected.

We then looked into correctly sampling and representing a field under iterative constraints and found a number of issues which needed addressing. Chiefly, we must make modifications to ensure that the sampled intensity field that we use for analysis has at least twice the bandwidth of the underlying amplitude field, this ensures that the pattern is well-sampled. We also identified the need to potentially limit the bandwidth of the field by a filter in order to ensure that the resultant amplitude pattern is implementable on our real modulator device without significant aliasing error.

In addition to this, we found that when using a sample rate higher than the pitch of the modulator by a factor of two, applying a pixelation routine as part of the phase only constraint (i.e. more closely matching the implemented modulation

on the SLM) results in an improved noise pattern. This was demonstrated for a square image pattern to result in a reduction of the COV of the high region by 55%. Later in chapter 7 we also observe that applying both filter and pixelate methodologies reduces the energy in a section of a problematic overlapping first order image by 20%.

We went on to investigate examples where we varied some of the parameters of the optical simulation to gain an understanding of their effect on image quality. Firstly, the size of the hologram was investigated when using a filter to limit the bandwidth of the hologram and a simple equation was proposed as an ideal limit on hologram size. Secondly, we showed an example of varying feature size below and above a filtered limit. This shows that getting feature sizes larger than the PSF of the optical system is unreliable beyond a limit (in this case 2.5 times the width of the PSF), and devolves into our usual noise pattern. Lastly, we looked at an example which varies image background value and observed that at zero background value a steep decline in achieved contrast is present. This is interesting to note and is most likely related to the reduction in the degree of freedom afforded the iterative system.

Chapter 5 then goes on to examine an active averaging method [53,54] proposed then discusses an alternative in coherence control of the system. At the end of Chapter 5 we gave an example of a simulated bus of lines pattern which had been iteratively optimised and compared it to its experimental reconstruction. Here we found good agreement, though with some increase in background noise.

Chapter 6 generalised the planar iterative algorithm by considering multiple output planes. This has been approached before by [38,40,87], but none have previously been interested in producing a quality high enough for lithography. Subsequently an issue when applying partial constraints with our iterative algorithm was discovered. This issue, that energy may pass through a partially constrained system without being attenuated, can be accounted for by ensuring that our phase-only constraint does not significantly increase the energy in our system.

We went on to investigate the minimum separation of the constraint image planes required to form a usefully high contrast image pattern through three experiments. Firstly, we assessed the separation and intensity pattern of two analytical zone plate point images. Unsurprisingly we found a significant drop off in intensity when points are separated by distances larger than $1\times$ the DOF of the system. Secondly, we took a constraint pattern consisting of two image plane constraints and separating them until contrast dropped below a useful limit. We found that the roll off of measured contrast with plane separations is slower than that of the two-point analytical approximation. This suggests lateral reinforcement from the image pattern has a

significant effect. Lastly, we investigated making a tilted line pattern out of many constraint planes, varying the inter-plane separation to find a minimum limit. In each of these cases we find that increasing complexity increases the required density of partial constraints, i.e. requires an increase in the number of constraint planes. We compare an analytically seeded iteration with a plane phase seeded approach and find some slight variances in noise pattern, and a significant difference in the contrast achieved with and without an analytical seed pattern at larger constraint plane separations.

In Chapter 7 we took several potentially useful implementation cases for the iterative algorithm and implemented them, exploring the issues and constraints in each case. As before we find that the more complex 'dog-leg' lines bus pattern works well and could potentially solve interconnect problems where anisotropically-etched silicon or ceramic or plastic packaging requires interconnect between upper and lower surfaces at high resolution.

Also implemented was a test antenna pattern with a much larger feature size. This was done using an averaged exposure of many noise-filled hologram images. We found that the pattern is implementable but that the constraint edge regions are sensitive at a much smaller depth of focus than would be expected from the larger feature size. For this particular pattern a single focal image at the average focal plane is more suitable because of the issues we find at the inter-constraint boundaries.

Finally, we looked at a new method allowing exposure of vertical surfaces by augmenting the exposure apparatus with a grating allowing 'non-flat' illumination of a substrate. The same transform which is applied for a simple line hologram is applied to the more complex crest patterns. Once again we found that the pattern works well and results in a good vertical exposure (despite an increase in noise). It is proposed that this kind of exposure technique could be applied in conjunction with the iterative bus design to implement high density, high resolution interconnect patterns on the edge of a silicon chip in place of wire bonding.

8.2 Further Applications and Exploitation

8.2.1 Method Improvements and Further Exploration

Alongside the design and implementation of this method we have identified various points where further exploration can potentially improve the viability of the method and its limitations.

The core limitations of the method do not revolve around simulation and prop-

agation of the field in anything except a computational scale sense. The multi-step propagation method has proved viable despite the noted issues with an error increase because of the windowing process. Relatively recent research [83] suggests that implementing a filtered AS propagation might allow us to propagate the field without having to re-window the image. Another alternative may be to implement a method such as the RS-convolution [88] method to similarly remove the limit on propagation distance, and potentially reduce computational burden.

Throughout this document we have not addressed the issue of quantisation of the resulting phase pattern. Using the SLM this was unnecessary as the quantisation required was so low as to be almost indistinguishable from a continuous phase pattern (256 phase levels). In simple tests not explored in this report, we have experienced useful patterns from this method with quantisation levels as low as 4-8 phase levels in simulation. For fixed masks the effect of this quantisation would need to be outlined in greater detail and some characterisations of lower limits on the number of phase levels implemented. Furthermore, some iterative schemes exist [46, 50] in which the quantisation states of the field can be iteratively introduced, which potentially results in a more optimised pattern. More complex modualtion technologies, for example manufacturing techniques that could easily produce phase-and-amplitude modulations or active modulators (or modulator pairs), will potentially change the landscape of what is achieveble in terms of holographic imaging systems. The potential existance of these technologies in the future however does not negate the usefulness of striving for a phase-only optimised solution, at this is the only case that will maximise energy throughput.

We have briefly mentioned the notion of introducing partially coherent sources as part of the source optic in this report 5.12. This can be done by reducing the coherence of a laser or by increasing the coherence of an incoherent source such as an arc-lamp. Though not discussed in detail in this thesis we have found that very small source patterns (i.e. high coherence) are required to maintain pattern contrast at large distances from the diffractive mask optic. Nonetheless, this process has potentially large applications even in 2D exposure cases where a large projection gap needs to be mitigated. In forthcoming literature we will publish our results describing our iteratively optimised attempts to get dense $10\mu m$ spot features at a $500\mu m$ offset using a fixed phase mask and arc-lamp illumination system. The result of this work once again moves us towards the integration of this holographic 3D design technique into a fixed mask mask alignment system. The largest issue with this approach is likely to be maintaining a high energy density in a source pattern which may have a relatively small source size due to a pinhole filter or

similar.

8.2.2 Further Applications

We have already discussed three potential lithographic applications: Sloped interconnect, vertical interconnect and antenna patterning. In addition to these, personal correspondence has indicated that this research might be useful in areas such as biological tissue stimulation where specific regions of non-uniformly shaped samples and specific sets of cells are stimulated by increased light intensity.

As noted briefly, some interaction has taken place with Suss micro-optic. Suss are interested in using holograms to overcome resolution limits for larger projection gaps as this is a common issue in proximity projection. Ongoing interaction will hopefully result in design of iterative holograms to overcome this issue.

8.3 Concluding Remarks

Through this thesis an iterative methodology for designing holograms has been developed which, through various modifications, allows useful lithographic exposure onto 3D substrates. Like all good engineering research the aim must be a practical system which would, we hope, have real benefits in industry. This project has brought us close to realising that aim. Where possible we have demonstrated and quantified the limitations of iteratively optimised phase-only holograms in terms of image and system parameters and applied this to 2D and then 3D methodologies. This resulted in a quantitative understanding of the parameters required to produce a continuous pattern on a 3D surface, which has been practically demonstrated repeatedly in the final chapters. Though practical boundaries still exist to an implementation of 3D holographic lithography as discussed above, the proposed system is eminently feasible and within current manufacturing capabilities. In conjunction with the design methodology of this report, we can safely say that almost any 3D surface geometry (within angular slope limitations) can be patterned by holographic lithography with simple line patterns, and by using an active modulator or by coherence tuning the same is true for almost any arbitrary shape. We furthermore demonstrate the extreme geometries such as the vertical wall patterns are also possible. Engineering challenges in alignment and system construction still exist, but pose no obvious barrier to the success of such a technique. It is our hope that further work developed upon this research will resolve these issues and see the approach implemented, allowing new classes of devices to become readily manufacturable.

Bibliography

[1] F. Dill. Optical Lithography. *IEEE Transactions on Electron Devices*, 22(7):440–452, July 1975.

[2] C. Harper. *Electronic Assembly Fabrication*. Mc Graw Hill, 2002.

[3] D. Gabor. A New Microscopic Principle. *Nature*, 161(4098):777–778, May 1948.

[4] M. Madou. *From MEMS to Bio-MEMS and Bio-NEMS*. Fundamentals of Microfabrication and Nanotechnology. CRC Press, third edition, 2012.

[5] Intel. Intel 22nm 3D Tri-Gate Transistor Technology. `http://newsroom.intel.com/docs/DOC-2032`, 2014. [Online; accessed May-2014].

[6] A. Wong. *Resolution Enhancement Techniques in Optical Lithography*, volume TT47 of *Tutorial Texts in Optical Engineering*. SPIE Press, 2001.

[7] P. Kersten, S. Bouwstra, and J. Petersen. Photolithography on micromachined 3D surfaces using electrodepositable photoresists. *Sensors and Actuators A*, 51:51–54, 1995.

[8] M. Madou. *From MEMS to Bio-MEMS and Bio-NEMS*. Manufacturing Techniques for Microfabrication and Nanotechnology. CRC Press, third edition, 2012.

[9] J. Cowling, R. McWilliam, A. Purvis, F. Soulard, P. Ivey, L. Seed, G. Williams, and J. Toriz-Garcia. Improvements in or relating to holography, WO201111413: 15 March 2011.

[10] J. Cowling, G. Williams, A. Purvis, R. McWilliam, J. Toriz-Garcia, N. Seed, F. Soulard, and P. Ivey. Three-dimensional holographic lithography by an iterative algorithm. *Optics letters*, 36(13):2495–2497, 2011.

[11] J. Toriz-Garcia, G. Williams, R. McWilliam, N. Seed, A. Purvis, F. Soulard, J. Cowling, and P. Ivey. Vertical tracks on the sidewall of a silicon die using 3d holographic photolithography. *Journal of Micromechanics and Microengineering*, 21(8):085034, 2011.

[12] J. Cowling, J. Toriz-Garcia, A. Purvis, R. McWilliam, G. Williams, F. Soulard, N. Seed, and P. Ivey. Time-division multiplexing of iterated holograms for lithography on 3d surfaces. *Digital Holography and Three-Dimensional Imaging*, 2012. (pp. DTu2C-5)

[13] J. Cowling, J. Toriz-Garcia, G. Williams, A. Purvis, R. McWilliam, F. Soulard, N. Seed, P. Ivey, and D. Claus. Holographic lithography on vertical surfaces. *Digital Holography and Three-Dimensional Imaging*, 2013. (pp. DW2A-3)

[14] J. Toriz-Garcia, J. Cowling, G. Williams, Q. Bai, N. Seed, A. Tennant, R. McWilliam, A. Purvis, F. Soulard, and P. Ivey. Fabrication of a 3d electrically small antenna using holographic photolithography. *Journal of Micromechanics and Microengineering*, 23(5):055010, 2013.

[15] J. Goodman. *Introduction to Fourier Optics*. Roberts and Company, third edition, 2005.

[16] J. Cooley and J. Tukey. An algorithm for the machine calculation of complex fourier series. *Math. comput*, 19(90):297–301, 1965.

[17] W. Singer, M. Totzeck, and H. Gross. *Handbook of Optical Systems*, volume 2. WILEY-VCH, 2005.

[18] R. Piestun and J. Shamir. Synthesis of Three-Dimensional Light Fields and Applications. *Proceedings of the IEEE*, 90(2):222–243, Feb 2002.

[19] G. Whyte and J. Courtial. Experimental demonstration of holographic three-dimenstional light shaping using a gerchberg-saxton algorithm. *New J. Phys.*, 7(117), May 2005.

[20] S. Maciej. Light propagation in the Fresnel region. New numerical approach. *J. Optical Communications*, 116:43–48, April 1995.

[21] W. Ash. Quantative charecterisation of cellular adhesions with total internal reflection holographic microscopy. Topical Meeting on Digital Holography and Three-Dimensional Imaging. OSA, April 2010.

[22] T. Tommasi and B. Bianco. Computer-generated holograms of tilted planes by a spatial frequency approach. *J. Opt. Soc. Am. A.*, 10(2):299–305, July 1993.

[23] D. Leseberg and C. Frere. Computer-Generated holograms of 3-D objects composed of tilted planar segments. *Applied Optics*, 27(14):3020–3024, July 1988.

[24] N. Delen and B. Hooker. Free-space beam propagation between arbitrarily oriented planes based on full diffraction theory: a fast fourier transform approach. *JOSA A*, 15(4):857–867, 1998.

[25] L. Onural. Exact solution for scalar diffraction between tilted and translated planes using impulse functions over a surface. *JOSA A*, 28(3):290–295, 2011.

[26] A. Maiden. *Holographic Lithography*. PhD thesis, School of Engineering, University of Durham, 2005.

[27] C. Frere, D. Leseberg, and O. Bryngdahl. Computer-generated holograms of three-dimensional objects composed of line segments. *J. Opt. Soc. Am. A*, 3(5):726–730, May 1986.

[28] C. Frere and O. Bryngdahl. Computer-generated holograms: reconstructions of curves in 3-D. *Optics Communications*, 60(6):369–372, Sep 1986.

[29] A. Maiden. High fidelity microlithography patterning using computer generated holograms. *Optics Letters*, 30:1300–1302, 2005.

[30] G. Lazarev, A. Hermerschmidt, S. Krüger, and S. Osten. Lcos spatial light modulators: Trends and applications. *Optical Imaging and Metrology: Advanced Technologies, W. Osten and N. Reingand editors, Wiley-VCH Verlag*, 2012.

[31] R. McWilliam, J. Cowling, G. Williams, A. Purvis, J. Toriz-Garcia, N. Seed, F. Soulard, and P. Ivey. High contrast imaging for photolithography using quantized computer generated holograms. *Applied Optics*, 2014. In Preparation.

[32] G. Williams, N. Seed, A. Purvis, A. Maiden, R. McWilliam, P. Ivey, and S. Johnson. NonPlanar interconnect. *Circuit World*, 31(2):1014, 2005.

[33] X. Ma and G. Arce. *Computational Lithography*, volume 77. Wiley. com, 2011.

[34] J. Fienup and C. Wackerman. Phase-retrieval stagnation problems and solutions. *JOSA A*, 3(11):1897–1907, 1986.

[35] R. Gerchberg and W. Saxton. Practical Algorithm for Determination of Phase from Image and Diffraction Plane Pictures. *Optik*, 35(2):237, 1972.

[36] J. Fienup. Invariant error metrics for image reconstruction. *Applied Optics*, 36(32):8352–8357, Nov 1997.

[37] R. Dorche, A. Lohmann, and Sinzinger. Fresnel ping-pong algorithm for two-plane computer-generated hologram display. *Applied Optics*, 33(5):869–875, February 1994.

[38] M. Makowski, M. Sypek, A. Kolodziejczyk, and M. Grzegorz. Three-plane phase-only computer hologram generated with iterative Fresnel algorithm. *Optical Engineering*, 44(12), Dec 2005.

[39] M. Makowski, M. Sypek, A. Kolodziejczyk, G. Mikuła, and J. Suszek. Iterative design of multiplane holograms: experiments and applications. *Optical Engineering*, 46(4):045802–045802, 2007.

[40] J. Xia and H. Yin. Three-dimensional light modulation using phase-only spatial light modulator. *Optical Engineering*, 48(2), Feb 2009.

[41] J. Fienup. Phase retrieval algorithms: a comparison. *Applied Optics*, 21(15):2758–2769, Aug 1982.

[42] J. Fienup. Reconstruction of an object from the modulus of its fourier transform. *Optics Letters*, 3(1):27–29, July 1978.

[43] G. Sinclair. Interactive application in holographic optical tweezers of a multi-plane Gerchberg-Saxton algorithm for three-dimensional light shaping. *Opt. Ex*, 12(8):1665–1670, April 2004.

[44] M. Makowski, M. Sypek, and A. Kolodziejczyk. Colourful reconstructions from a thin multi-plane phase hologram. *Opt. Ex*, 16(15), Jul 2008.

[45] M. Makowski, M. Sypek, I. Ducin, A. Fajst, A. Siemion, J. Suzek, and A. Kolodziejczyk. Experimental evaluation of a full-color compact lensless holographic display. *Opt. Ex*, 17(23), Oct 2009.

[46] F. Wyrowski. Iterative Quantization of digital amplitude holograms. *Applied Optics*, 28(18):3864–3870, Sep 1989.

[47] F. Bernhardt, F. Wyrowski, and O. Bryngdahl. Coding and binarisation in digital fresnel holography. *Optics Communications*, 77(1):4–8, June 1990.

[48] T. Peter, F. Wyrowski, and O. Bryngdahl. Importance of initial distribution for iterative calculation of quantised diffractive elements. *J. Modern Optics*, 40(4):591–600, June 1993.

[49] F. Wyrowski. Diffraction efficiency of analog and quantised digital amplitude holograms: analysis and manipulation. *J. Opt. Soc. Am. A.*, 7(3):383–393, March 1990.

[50] M. Skeren, I. Richter, and P. Fiala. Iterative Fourier transform algorithm: comparison of various approaches. *J. Mod. Opt.*, 49(11):1851–1870, Dec 2002.

[51] P. Senthilkumaran, F. Wyrowski, and H. Schimmel. Vortex stagnation problem in iterative fourier transform algorithms. *Optics and lasers in engineering*, 43(1):43–56, 2005.

[52] F. Wyrowski and O. Bryngdahl. Iterative fourier-transform algorithm applied to computer holography. *JOSA A*, 5(7):1058–1065, 1988.

[53] J. Amako, H. Miura, and T. Sonehara. Speckle-noise reduction on kinoform reconstruction using a phase-only spatial light modulator. *Applied optics*, 34(17):3165–3171, 1995.

[54] C. Bay, N. Hubner, J. Freeman, and T. Wilkinson. Maskless photolithography via holographic optical projection. *Optics Letters*, 35(13):2230–2232, July 2010.

[55] L. Golan and S. Shoham. Speckle elimination using shift-averaging in high-rate holographic projection. *Opt. Ex*, 17(3):1330–1339, 2009.

[56] M. Pasienski and B. DeMarco. A high-accuracy algorithm for designing arbitrary holographic atom traps. *arXiv preprint arXiv:0712.0794*, 2007.

[57] A. L Gaunt and Z. Hadzibabic. Robust digital holography for ultracold atom trapping. *Scientific reports*, 2, 2012.

[58] R. Voelkel, U. Vogler, A. Bich, P. Pernet, K. Weible, M. Hornung, R. Zoberbier, E. Cullmann, L. Stuerzebecher, and T. Harzendorf. Advanced mask aligner lithography: new illumination system. *Opt. Ex*, 18(20):20968–20978, 2010.

[59] M. Seldowitz, J. Allebach, and D. Sweeney. Synthesis of digital holograms by direct binary search. *Applied Optics*, 26(14):2788–2798, Jul 1987.

[60] B. Jennison and J. Allebach. Efficient design of dirct-binary-search computer-generated holograms. *J. Opt. Soc. Am. A.*, 8(4):652–660, Dec 1991.

[61] F. Wyrowski, E. Kley, S. Buhling, A. Nellissen, L. Wang, and M. Dirkzwager. Proximity printing by wave-optically designed masks. In Frank Wyrowski, editor, *Wave-Optical Systems Engineering*, volume 44 of *Proceedings of SPIE*, pages 130–139, 2001.

[62] V. Singh. Making a trillion pixels dance. In H. Levinson, editor, *Optical Microlithography XXI*, volume 6924, 2008.

[63] F. Soulard, A. Purvis, R. McWilliam, J. Cowling, G. Williams, Toriz-Garcia J., N. Seed, and Ivey. P. Iterative zero-order suppression from an off-axis hologram based on the 2d hilbert transform. In *Biomedical Optics and 3-D Imaging*, page DSu3C.4. Optical Society of America, 2012.

[64] B. Brown and A. Lohmann. Computer-generated binary holograms. *IBM Journal of research and Development*, 13(2):160–168, 1969.

[65] J. Bucklew and N. Gallagher. Comprehensive error models and a comparative study of some detour-phase holograms. *Applied Optics*, 18(16):2861–2869, 1979.

[66] G. Makey, M. El-Daher, and K. Al-Shufi. Utilization of a liquid crystal spatial light modulator in a gray scale detour phase method for fourier holograms. *Applied optics*, 51(32):7877–7882, 2012.

[67] ALPAO. Astronomical Adaptive Optics. `http://www.alpao.com/Applications/Astronomy.htm`, 2013. [Online; accessed Nov-2013].

[68] Texas Instruments. Binary digital micromirror devices (DMD). `http://www.ti.com/ww/en/analog/mems/dlplightcommander/index.shtml?DCMP=DLP_Light_Commander&HQS=Other+OT+dlplightcommander`, 2013. [Online; accessed Nov-2013].

[69] Holoeye. Liquid Crystal on Silicon Spatial Light Modulators. `http://holoeye.com/spatial-light-modulators/`, 2013. [Online; accessed Nov-2013].

[70] E. Van Putten, I. Vellekoop, and A. Mosk. Spatial amplitude and phase modulation using commercial twisted nematic lcds. *Applied optics*, 47(12):2076–2081, 2008.

[71] S. Zwick, T. Haist, M. Warber, and W. Osten. Dynamic holography using pixelated light modulators. *Applied optics*, 49(25):F47–F58, 2010.

[72] K. Chellappan, E. Erden, and H. Urey. Laser based displays a review. *J. Applied Optics*, 49(25), September 2010.

[73] F. Yaras, H. Kang, and L. Onural. State of the art in holographic displays a survey. *J. Display Technology*, 6(10), October 2010.

[74] R. Tudela, E. Martin-Badosa, I. Labastida, S. Vallmitjana, I. Juvells, and A. Carnicer. Full complex fresnel holograms displayed on liquid crystal devices. *Journal of Optics A: Pure and Applied Optics*, 5(5):S189, 2003.

[75] J. Herrera-Fernandez and L. Sanchez-Brea. Double diffractive optical element system for near-field shaping. *Applied Optics*, 50(23):4587–4593, 2011.

[76] Judit Remenyi, Peter Varhegyi, László Domján, Pál Koppa, and Emõke Lõrincz. Amplitude, phase, and hybrid ternary modulation modes of a twisted-nematic liquid-crystal display at˜ 400 nm. *Applied optics*, 42(17):3428–3434, 2003.

[77] R. Voelkel, U. Vogler, A. Bramati, T. Weichelt, L. Stuerzebecher, U. Zeitner, K. Motzek, A. Erdmann, M. Hornung, and R. Zoberbier. Advanced mask aligner lithography (amalith). In *SPIE Advanced Lithography*, pages 83261Y–83261Y. International Society for Optics and Photonics, 2012.

[78] U. Vogler, F. Windrich, A. Schenke, R. Volkel, M. Bottcher, and R. Zoberbier. Cost-effective lithography for tsv-structures. In *Electronic Components and Technology Conference (ECTC), 2011 IEEE 61st*, pages 1407–1411. IEEE, 2011.

[79] H. Aagedal, M. Schmid, T. Beth, S. Teiwes, and F. Wyrowski. Theory of speckles in diffractive optics and its application to beam shaping. *Journal of modern optics*, 43(7):1409–1421, 1996.

[80] J. Goodman. *Speckle phenomena in optics: theory and applications*. Roberts and Company Publishers, 2007.

[81] F. Wyrowski. Diffractive optical elements: iterative calculation of quantized, blazed phase structures. *JOSA A*, 7(6):961–969, 1990.

[82] M. Sypek and M. GoÂ. Image multiplying and high-frequency oscillations effects in the fresnel region light propagation simulation. *Optical Engineering*, 42(11):3158–3164, 2003.

[83] Kyoji Matsushima and Tomoyoshi Shimobaba. Band-Limited Angular Spectrum Method for Numerical Simulation of Free-Space Propagation in Far and Near Fields. *Opt. Ex*, 17(22):19662–19673, OCT 26 2009.

[84] Sarantel. Sarantel, Antenna Manufacturer, 3D Lithography. `http://en.wikipedia.org/wiki/Sarantel`, 2014. [Online; accessed Sep-2014].

[85] T. Nellissen, L. Wang, and R. Wehrens. A novel photolithographic method for realising 3D interconnection patterns on electronic modules. In *Microelectronics and Packaging Conference*, pages 347–351, June 2003.

[86] T. Nellissen, M. Botermans, M. Burghoorn, J. Van Delft, E. Van Grunsven, J. Scheer, and M. De Samber. Development of an advanced three-dimensional mcm-d substrate level patterning technique. 2004.

[87] J. Waters. Three-dimensional fourier-transform method for synthesizing binary holograms. *J. Optical Society of America*, 58(9), September 1968.

[88] F. Shen and Wang A. Fast-fourier-transform based numerical integration method for the rayleigh-sommerfeld diffraction formula. *Appl. Opt.*, 45(6):1102–1110, Feb 2006.

Appendix A

Code

A.1 Error Calculation

The SSE calculation is simply performed on the field and ideal image amplitudes as follows:

```
% Normalised SSE for amplitude of two images
function output = SSE(ideal,actual)

    sq_error = ( abs( abs(actual) - abs(ideal) ) .^2 );
    sq_ideal = ( abs(ideal).^2 );
    output   = ( sum2D(sq_error) ) / ( sum2D(sq_ideal) );

end
```

However this value is variant depending upon the scale of the image intensity. This can be minimised using the following routine:

```
% Get minimum SSE for scale invariant intensity.
function [error]=min_SSE(ideal,actual,accuracy)

    % Initialise Values
    scale_fac = 1;
    step = 100;

    % Calculate First Error
    lasterror = SSE(ideal,actual.*scale_fac);

    % Loop until accuracy satisfies input constraint
    while(abs(step)>accuracy)

        scale_fac = scale_fac + step;

```

```
16          % calculate error
17          nexterror=SSE(ideal,actual.*scale_fac);
18
19          %refine accuracy
20          if (nexterror > lasterror)
21          step = -step/10;
22          end
23
24          lasterror = nexterror;
25      end
26
27      error = lasterror;
28
29  end
```

A.2 Transform Calculation

```
1  %% Angular-spectrum transfer function Calculation
2  function [output] = ...
       calc_AS_TF(z,lambda,x_samples,y_samples,sample_pitch)
3      vx=(-x_samples/2:x_samples/2-1).*(1/(x_samples*sample_pitch));
4      vy=(-y_samples/2:y_samples/2-1).*(1/(y_samples*sample_pitch));
5
6      [VY,VX] = meshgrid(vy,vx);
7
8      output=
9      fftshift(
10         rot90(exp(2*pi*1i*z.*sqrt(((1/lambda)^2-VX.^2-VY.^2))))
11     );
12 end
```

A.3 Multi-plane Partially Constrained Iteration Example

Example code for a partially constrained multi-plane iteration.

```
1  % Multi-Plane Iteration Example
2  % initialisations
3  clear;
4  clc;
5
6  n_iter = 50; % number of iterations
```

```
lam = 405e-9; % wavelength
sp = 2e-6; % sample pitch
z3 = 0.08; % modulator to image distance

dz = 7.583969990505682e-04; % inter-plane separation

% hologram size variables
hologram_width = 512 * 8e-6;
hologram_height = 512 * 8e-6;

% image depth
z2 = hologram_width * tan(pi/4);

% simulation samples
xs = (hologram_width/sp)*2
ys = (hologram_height/sp)*2;

% hologram size in samples
sxs = hologram_width/sp;
sys = hologram_height/sp;

% number of constraint planes
n_planes = ceil( z2./dz ) + 1;

% image sizes corrected for extra iteration
realz2 = dz * (n_planes-1);
z2diff = realz2-z2;
realz3 = z3 - z2diff/2;
realz1 = realz2 + realz3;

% bandwidth limit as a proportion of max bandwidth
filterprop = 0.25;

% resolutions calculated based on system parameters
NA = (lam * filterprop) / (2 * sp) ;
theta = asin(NA);
xBW = 2*sin(theta)/lam;
xRes = 1/xBW
zBW = 1/lam - cos(theta)./lam;
zSpot = 1./zBW

% maximum z propagation distance in one step
mz=((xs)*sp^2)/(2*lam)/2

% number of steps determined by max z above
```

```
52 n_step=round( realz1/mz )
53
54 % maximum bandwidth filter mask
55 mask=fftshift(gen_bl_mask_elip(xs,ys,1/(xs/2),1/(ys/2),filterprop));
56
57 % calculation of filtered transfer functions
58 H_dz   = calc_AS_TF_GPU(-dz,lam,xs,ys,sp).*mask;
59 H_z3md = calc_AS_TF_GPU((-realz3+dz)/n_step,lam,xs,ys,sp).*mask;
60 H_z1   = calc_AS_TF_GPU( realz1/n_step ,lam,xs,ys,sp).*mask;
61
62 %% example surface calculation ( slope )
63 if n_planes>2
64     [X,Y] = meshgrid(1:sxs,1:sys);
65
66     X=X-sxs/2;
67     Y=Y-sys/2;
68     SURF=X;
69     SURF=SURF-min2d(SURF);
70
71     % surf mask used to seperate parital constraints
72     surf_mask = gen_surf_mask(SURF,n_planes);
73 else
74     if n_planes==2
75         surf_mask=zeros(sys,sxs);
76         surf_mask(:,sxs/2:end)=1;
77     else
78         surf_mask=zeros(sxs,sys);
79     end
80 end
81
82 surf_mask=int16(surf_mask);
83
84 %% example image calculation
85
86 % line image size parameters
87 linepitch = 16; %dims in pixels
88 linewidth = 1;
89
90 % size limits
91 start = 0.25; %proportions of xs/ys to start/stop line
92 stop = 0.75;
93
94 %draw lines on image mask
95 for m = 1: nlines
96     for n=1:linewidth
```

```
97          linestart(n+(m-1)*linewidth,2) = sxs*start;
98          linestart(n+(m-1)*linewidth,1) = ...
                fix(sys/2)-fix(linewidth/2)+n+linepitch* ...
                (m-1)-fix(nlines/2)*linepitch;
99
100         linestop(n+(m-1)*linewidth,2) = sxs*stop;
101         linestop(n+(m-1)*linewidth,1) = ...
                fix(sys/2)-fix(linewidth/2)+n+linepitch* ...
                (m-1)-fix(nlines/2)*linepitch;
102     end
103 end
104 image = zeros(sxs,sys);
105 for line = 1:nlines*linewidth
106     % use bresenham algorithm to draw pixelated line.
107     image = bresenham( image , [linestart(line,:) ; ...
            linestop(line,:)] , 0 , 1 );
108 end
109
110 % set image background value
111 image = image*0.975+0.025;
112 image = pad(image,xs,ys,0);
113
114 %% set up initial hologram
115
116 % create initial functions
117 HOLO = exp( 1i .* ones( ys , xs ) .* 2 * pi );
118 % limit bandwidth
119 HOLO = simple_bandlimit_circ_spectrum( HOLO , 0.25 );
120 % correct to phase only
121 HOLO=HOLO./abs(HOLO);
122 % limit spatial region
123 HOLO=bandlimit(HOLO,sxs,sys,0);
124
125 % initial energy correction
126 % maintain sum of squares in image/hologram
127 eh = calc_energy(HOLO);
128 eim = calc_energy( image );
129 e_h_factor=eim/eh;
130 HOLO=HOLO.*sqrt(e_h_factor);
131
132 % propagate initial hologram
133 P=convolve_with_TF_N_GPU(HOLO,H_z1,n_step,sxs,sys);
134
135 %% iteration loop
136 for n=1:n_iter
```

```
137     n
138     tic
139
140     %% loop over constraint planes
141     for m=1:n_planes
142         m
143         % constraints
144         P=mask_enforce(P,surf_mask,image,m);
145
146         % Δ backward transform
147         P=convolve_with_TF_N_GPU(P,H_dz,1,sxs,sys);
148
149     % end of planes loop
150     end
151
152     %% (z3—Δ) backwards transform
153     P=convolve_with_TF_N_GPU(P,H_z3md,n_step,sxs,sys);
154
155     % diffraction screen constraints
156     % phase—only with energy conservation
157     e=calc_energy(P);
158     P=P./abs(P);
159     e_new=calc_energy(P);
160     e_factor=e/e_new;
161     P=P.*sqrt(e_factor);
162
163     % save hologram
164     HOLO=P;
165
166     % forward propagation
167     P=convolve_with_TF_N_GPU(P,H_z1,n_step,sxs,sys);
168
169     toc*(n_iter—n)
170
171 % end iteration loop
172 end
```

A.4 Auxillary Scripts

Apply circular mask to field spectrum at a given proportion of maximum bandwidth.

```
1 function [output] = simple_bandlimit_circ_spectrum(input,prop)
2
3 [XS,YS]=size(input);
```

```
4
5  FT=fftshift(fft2(input));
6
7  mask=gen_bl_mask_elip(XS,YS,1/(max2d(XS)),1/(max2d(YS)),prop);
8
9  FT=FT.*mask;
10
11 output = ifft2(ifftshift(FT));
```

Limit the spatial size of a field

```
1  function [output] = bandlimit(input,xlim,ylim,padval)
2
3  [XS,YS]=size(input);
4
5  input(1:uint32(XS/2-(xlim/2)),:) = padval;
6  input(uint32(XS/2+(xlim/2)):end,:) = padval;
7  input(:,1:uint32(YS/2-(ylim/2))) = padval;
8  input(:,uint32(YS/2+(ylim/2)):end) = padval;
9
10 output = input;
```

Calculate energy of amplitude field (sum of squares)

```
1  %% Calc Energy
2  % Calc energy of wave field by summing the square of the field ...
       values.
3
4  function output=calc_energy(input)
5      input=abs(input).^2;
6      input=sum(sum(input));
7      output=input;
8  end
```

Convolve with transfer function repeatedly to repeat propagation at maximum step distance.

```
1  %% convolve with TF N times
2  function output = convolve_with_TF_N(A,H,N,blx,bly)
3
4      for a=1:N
5          A=convolve_with_TF(bandlimit(A,blx,bly,0),H);
6      end
7
8      output=A;
9
10 end
```

Apply region of quantised mask to image, leaving unconstrained regions unchanged.

```
%% Mask Enforce
function [op_plane] = mask_enforce(ip_plane,surf_mask,image,level)
    [ys,xs] = size(ip_plane);

    mask = rot90(logical(pad(double(rot90(surf_mask,3)) ...
        -level+1,xs,ys,0))); % area not to touch

    amp_prev=abs(ip_plane); % old amplitude

    amp_mask = amp_prev.*mask + image.*~mask; % new amplitude, ...
        constrained in some areas same as prev in others

    %op_plane=exp(1i.*angle(ip_plane)).*amp_mask;
    op_plane=(ip_plane./abs(ip_plane)).*amp_mask; % applied to phase
end
```